MATB24 Lecture Notes
Linear Algebra

Pourya

Contents

1 Field

1.1 Preliminary

In the Mathematical Universe everything starts off with sets. A set X is a collection of mathematically defined objects.

Definition 1.1.1 (Cartesian Products)**.** Given two sets X and Y, the **Cartesian Product** of these two sets, denoted by $X \times Y$, is the set $\{(a, b) : \ a \in X, \ b \in Y\}$

Remark 1.1.1. There is no restriction on the size of these sets. X and Y can be both finite, infinite, or one finite and infinite. And we denote the size of a set by $|X|$. Even though there are different "types" of infinity, we only care if a set is not finite, then its size is infinite, loosely speaking.

Definition 1.1.2 (Binary Operation)**.** Given a set X, a **Binary Operation** $\star$ on X is a function defined on the Cartesian product of X with itself, to X. That is

$$\begin{aligned} \star :& X \times X \to X \\ &(a, b) \longmapsto \star\big((a, b)\big) = a \star b \end{aligned}$$

Where $\star((a, b)) = a \star b$ is called the **image** of (a, b) under $\star$. Note that assigning the ordered pair (a, b) to an element in X is equivalent to saying that X is **closed** under the operation $\star$.

Remark 1.1.2. A binary operation $\star$ on a set X does not have to be either *injective* and/or *surjective.*

Example 1.1.1. Consider the set $X = \mathbb{Z}^+ = \{1, 2, ...\}$. Is $-$ (usual subtract) a binary operation in $\mathbb{Z}^+$?

Answer: No. Since for instance, $(1, 2) \longmapsto -\big((1, 2)\big) = 1 - 2 = -1 \notin \mathbb{Z}^+$

Definition 1.1.3 (Commutativity)**.** Let $\star$ be the binary operation on a set X. We say $\star$ is **commutative** if for all $a, b \in X$, we have that

$$a \star b = b \star a$$

Remark 1.1.3. Take the usual subtraction on the integers:

$$- : \mathbb{Z} \times \mathbb{Z} \to \mathbb{Z}$$

Now this is a binary operation on $\mathbb{Z}$ but it is not a **commutative** binary operation, since we observe that

$$-((3, 2)) = 3 - 2 = 1 \neq -1 = 2 - 3 = -((2, 3))$$

Definition 1.1.4 (Associativity)**.** We say the binary operation $\star$ is **associative** if for all $a, b, c \in S$

$$(a \star b) \star c = a \star (b \star c)$$

Remark 1.1.4. Note that subtraction is not **associative** on $\mathbb{Z}$ either, for $(2-3)-5 = -6 \neq 4 = 2-(3-5)$. However, addition is both a commutative and associative binary operation on $\mathbb{Z}$.

Remark 1.1.5. If the binary operation is associative we unambiguously write $a \star b \star c$.

Example 1.1.2. Consider the set $X = \mathbb{Q}\backslash\{0\}$ and the map $\star$ is depicted by $a \star b = \frac{a}{b}$. Is $\star$ a binary operation?

Answer: Yes. If $a, b \in \mathbb{Q}\backslash\{0\} \implies \exists\, m, n, s, t \in \mathbb{Z}$ with $m \neq 0$, $n \neq 0$ and $t \neq 0$, $s \neq 0 \ni$: $a = \frac{m}{n}$ and $b = \frac{s}{t}$. Then, $a \star b = \frac{a}{b} = \frac{m}{n} \times \frac{t}{s} = \frac{mt}{ns}$. Since they are all non-zero integers, their products are also a non-zero integers.

Remark 1.1.6. Recall that $\mathbb{Q} = \{\frac{a}{b} : \ a, b \in \mathbb{Z},\ b \neq 0 \ \wedge (a,b) = 1\}$ where $(a,b) = 1$ means that the greatest common divisor is 1, and hence called *relatively prime*

Exercise 1.1.1. Is the binary operation in 1.1.2 associative?

1.2 Fields

Motivation: The study of fields is concerned with sets possessing two binary operations, conventionally called addition '+' and multiplication '×', where these binary operations satisfy certain mathematical properties. Since sets are either finite or infinite we do have structures (fields) that are finite or infinite. But why or how did the mathematicians develop such structures? One can argue that it was enticed by and arose from trying to solve for the following quadratic equation: $x^2 + 1 = 0$. This mathematical structure allows us to define certain spaces over which the above quadratic equation has a solution. Other applications may revolve around the notion of the construction of geometrical objects using compass and straight edge. We may see an example if time permits.

Definition 1.2.1 (Field)**.** A field is a triple $(\mathbb{F}, +, \times)$ where F is the underlying set with $+$ and $\times$ as two binary operations satisfying the following axioms given $\forall a, b, c \in \mathbb{F}$.

The first four axioms are related to the structure $(\mathbb{F}, +)$, illustrated as follows:

(i) $a + b = b + a$,

(ii) $(a + b) + c = a + (b + c)$

(iii) $\exists\, z \in \mathbb{F}$ such that $\forall\, a \in F$, $z + a = a + z = a$, **(existence of additive identity)**

(iv) $\forall\, a \in \mathbb{F}\ \exists\, a' \in \mathbb{F}$ such that $a + a' = a' + a = z$, **(existence of additive inverse)**

The following axioms related to the structure $(\mathbb{F}, \times)$ are given as follows:

(I) $a \times b = b \times a$,

(II) $(a \times b) \times c = a \times (b \times c)$, i.e., associativity under multiplication.

(III) $\exists\, e \in \mathbb{F}$ such that $\forall\, a \in F$, $e \times a = a \times e = a$, **(existence of multiplicative identity)**

(IV) $\forall\, a \in \mathbb{F} \setminus \{z\}\ \exists\, i \in \mathbb{F}$ such that $a \times i = i \times a = e$, **(existence of multiplicative inverse)**

(V) $a \times (b + c) = a \times b + a \times c$ and $(b + c) \times a = b \times a + c \times a$ **(multiplication distributivity over +)**

Notation: From this point onward, we denote the elements $z,\ a',\ e,\ i$ in axioms (iii), (iv), (III), (IV) by $0_{\mathbb{F}},\ -a,\ 1_{\mathbb{F}},\ a^{-1}$, respectively. We used the index $\mathbb{F}$ to indicate that these elements are different depending on the structure of the fields. We omit such indices if it is clear from the context.

Theorem 1.2.1. *Let $(\mathbb{F}, +, .)$ be a Field. Then for any non-zero element of $\mathbb{F}$, $a \in \mathbb{F} \setminus \{0_F\}$, its multiplicative inverse, a^{-1} is unique.*

Proof. Let $a \in \mathbb{F} \setminus \{0_F\}$. Suppose there are two elements $b, c \in F$ such that $a.b = 1_F$ and $a.c = 1_F$.

$$\begin{aligned} b &= b.1_{\mathbb{F}} && \text{(multiplicative identity)} \\ &= b.(a.c) && \text{(multiplicative inverse)} \\ &= (b.a).c && \text{(multiplicative associativity)} \\ &= (1_{\mathbb{F}}).c && \text{(multiplicative inverse)} \\ &= c \end{aligned}$$

$\therefore b = c = a^{-1}$ ∎

Exercise 1.2.1. Prove that $0_{\mathbb{F}}$, $1_{\mathbb{F}}$ and additive inverse element $-a$ of any element $a \in \mathbb{F}$ are unique.

Example 1.2.1. $(\mathbb{Q}, +, .)$ and $(\mathbb{R}, +, .)$ are two examples of Fields where $+$ and $.$ are the usual addition and multiplications.

Answer: We can easily verify all the axioms are satisfied. Note $0_{\mathbb{Q}} = 0_{\mathbb{R}} = 0$ and $1_{\mathbb{Q}} = 1_{\mathbb{R}} = 1$. Furthermore, for any nonzero real number x, $x^{-1} = \frac{1}{x}$ is the corresponding multiplicative inverse.

Example 1.2.2. Consider the triple $(\mathbb{Z}_3, + \mod 3, . \mod 3)$. This is another example of a Field. It is just a FINITE Field.

Answer: First note that $\mathbb{Z}_3 = \{0, 1, 2\}$. Note $0_{\mathbb{Z}_3} = 0$ and $1_{\mathbb{Z}_3} = 1$.
Also, $2^{-1} = 2$ since $2.2 = 4 \equiv 1 \mod 3$. Also $-2 = 1$ and $-1 = 2$ since $2 + 1 = 3 \equiv 0 \mod 3$

Remark 1.2.1. Note that example 1.2.2 should convince us that given a Field $\mathbb{F}$ two elements of $\mathbb{F}$ can be the additive inverses of each other, or an element can be the multiplicative inverse of itself! Of course such relation can be swapped as well.

Remark 1.2.2. Recall that given two integers n and $s \in \mathbb{Z}_{>0}$, we say $n \equiv m \mod s \iff$ there is $q \in \mathbb{Z}$, $\ni: \quad n = sq + m$

Remark 1.2.3. If $a_1 \equiv b_1 \mod n$ and $a_2 \equiv b_2 \mod n$ then $a_1 + a_2 \equiv b_1 + b_2 \mod n$. That is because by definition we have integers p, q such that $a_1 = b_1 + pn$ and $a_2 = b_2 + qn$. Then $a_1 + a_2 = b_1 + b_2 + (p + q)n$. In a similar fashion we could show that $a_1 a_2 \equiv b_1 b_2 \mod n$

Proposition 1.2.1. *Let a and b be any two elements of a field $\mathbb{F}$. Then:*

1) $(-1_{\mathbb{F}})a = -a$

2) $(-a)b = a(-b) = -(ab)$

3) $(-a)(-b) = ab$

Proof.

1) Let a be any element in $\mathbb{F}$. To establish part 1), we need to show that the additive inverse of the element a is $(-1_{\mathbb{F}})a$. That is equivalent to showing $(-1_{\mathbb{F}})a + a = 0_{\mathbb{F}}$.
To this end,

$$\begin{aligned}
(-1_{\mathbb{F}})a + a &= (-1_{\mathbb{F}})a + 1_{\mathbb{F}}.a && \text{(multiplicative identity)}\\
&= ((-1_{\mathbb{F}}) + 1_{\mathbb{F}})a && \text{(multiplication distributivity over +)}\\
&= 0_{\mathbb{F}}a && \text{(additive inverse)}\\
&= 0_{\mathbb{F}}a + 0_{\mathbb{F}} && \text{(additive identity)}\\
&= 0a + \Big(0a + \big(-(0a)\big)\Big) && \text{(additive inverse)}\\
&= (0a + 0a) + \big(-(0a)\big) && \text{(Associativity of +)}\\
&= (0 + 0)a + \big(-(0a)\big) && \text{(multiplication distributivity over +)}\\
&= 0a + \big(-(0a)\big) && \text{(additive identity)}\\
&= 0_{\mathbb{F}} && \text{(additive inverse)}
\end{aligned}$$

2) First,

$$\begin{aligned}
(-a)b &= \big((-1).a\big)b && \text{(by part 1))}\\
&= (-1)(ab) && \text{(associativity of multiplication)}\\
&= -(ab) && \text{(by part 1))}
\end{aligned}$$

Now,

$$\begin{aligned}
a(-b) &= a\big((-1)b\big) && \text{(by part 1))}\\
&= \big(a(-1)\big)b && \text{(associativity of multiplication)}\\
&= \big((-1)a\big)b && \text{(commutativity of multiplication)}\\
&= (-a)b && \text{(by part 1))}
\end{aligned}$$

∎

Exercise 1.2.2. Prove part 3) of proposition 1.2.1

Remark 1.2.4. Note in the proof of part 1) of 1.2.1 we also proved that for every $a \in \mathbb{F}$, $0_{\mathbb{F}}.a = 0_{\mathbb{F}}$

Definition 1.2.2 (Idempotent Element)**.** In an "algebraic structure" $(\mathcal{R}, +, \cdot)$ weaker than the notion of the field, an element $\alpha \in \mathcal{R}$ is called **idempotent** if $\alpha^2 = \alpha.\alpha = \alpha$.

Example 1.2.3. Let $(\mathbb{F}, +, .)$ be a field. Characterize/identify all idempotent elements of $\mathbb{F}$

Answers: Let $\alpha \in \mathbb{F}$ be an idempotent element. Clearly if $\alpha = 0_{\mathbb{F}}$ we have that $0_F^2 = 0_{\mathbb{F}}.0_{\mathbb{F}} = 0_{\mathbb{F}}$
Hence, suppose $0_{\mathbb{F}} \neq \alpha$
By definition we have that $\alpha^2 = \alpha \quad (1)$.
Now since α is nonzero, it has a multiplicative inverse α^{-1}.
Therefore multiplying this multiplicative inverse on both sides of equation (1) we obtain that
$\alpha^{-1}(\alpha^2) = \alpha^{-1}\alpha = 1_{\mathbb{F}}$. But $\alpha^{-1}(\alpha^2) = (\alpha^{-1}\alpha)\alpha = 1_{\mathbb{F}}\alpha = \alpha$
$\therefore \alpha$ is either the multiplicative identity of $\mathbb{F}$, $1_{\mathbb{F}}$, or the additive identity of $\mathbb{F}$, $0_{\mathbb{F}}$.

Example 1.2.4. Consider the following algebraic structure

$$(M_{2\times 2}(\mathbb{R}), +, .),$$

where $M_{2\times 2}(\mathbb{R}) = \left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} : a, b, c, d \in \mathbb{R} \right\}$, and + and . are the usual matrix addition & and multiplication respectively. Identify some of the idempotent elements.

Answer: First note that $(M_{2\times 2}(\mathbb{R}), +, .)$ is not a field! Why not?
Consider the matrix $\hat{A} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$, which is neither $1_{M_{2\times 2}(\mathbb{R})}$ nor $0_{M_{2\times 2}(\mathbb{R})}$. (What is $1_{M_{2\times 2}(\mathbb{R})}$ and $0_{M_{2\times 2}(\mathbb{R})}$?)
$\hat{A}^2 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}^2 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} = \hat{A}$. Hence $\hat{A}$ is idempotent.
Consider $\hat{B} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$. Clearly $\hat{B}^2 = \hat{B}$ as well.

Exercise 1.2.3. Show $(\mathcal{M}_{2\times 2}(\mathbb{R}), +, .)$ is not a field.

Definition 1.2.3 (Algebraic Structure)**.** The triple $(\mathcal{R}, +, \cdot)$ is called an **algebraic structure** if the following are satisfied:

I) $+$ **is commutative**

II) $+$ **is associative**

III) $+$ **yields additive identity**

IV) $+$ **yields additive inverse**

i) $\cdot$ **is associative**

ii) $\cdot$ **is distributive over** $+$

Remark 1.2.5. Every field is an algebraic structure, but not every algebraic structure is a field. In some algebraic structure we might even have that $\cdot$ is commutative, or yields multiplicative identity.

$(\mathcal{R}, +, \cdot)$ **is mathematically known as a Ring.**

Proposition 1.2.2. *Let* $(\mathbb{F}, +, \cdot)$ *be a field. Given any two elements* $\alpha,\ \beta \in \mathbb{F}$ *we have* $\alpha\beta = 0_{\mathbb{F}} \iff \alpha = 0_{\mathbb{F}}$ *or* $\beta = 0_{\mathbb{F}}$

Proof. Remark 1.2.4 establishes that if $\alpha = 0$ or $\beta = 0$ then $\alpha\beta = 0$. Hence, we proceed to show that if $\alpha\beta = 0$, then either $\alpha = 0$ or $\beta = 0$. WLOG assume that $\alpha \neq 0$. Hence, the multiplicative inverse of α exists. Then $\alpha^{-1}(\alpha\beta) = \alpha^{-1}0 = (\alpha^{-1}\alpha)\beta = 1\beta = \beta = \alpha^{-1}0 = 0$ ∎

Remark 1.2.6. Note that 1.2.2 can only be employed to show that an algebraic structure $(R, +, \cdot)$ may not be a field. Consider the following example:

Example 1.2.5. Consider the algebraic structure $(\mathbb{Z}_6, + \mod 6, \cdot \mod 6)$.
Determine whether or not it constitutes a field.

Answer: By proposition 1.2.2, in a field, given any two elements $\alpha,\ \beta$ we have $\alpha\beta = 0 \iff \alpha = 0$ or $\beta = 0$. Since $\alpha = 0$ or $\beta = 0 \implies \alpha\beta = 0$ in any algebraic structure.
If we could find nonzero elements $\alpha,\ \beta$ in $\mathbb{Z}_6$ such that $\alpha\beta = 0$ then we can conclude that $(\mathbb{Z}_6, + \mod 6, \cdot \mod 6)$ is NOT a field. To this end, take $\alpha = 3$ and $\beta = 2$.
Clearly $\alpha\beta = 3 \cdot 2 \mod 6 = 0$. But neither α nor β is the additive identity of $\mathbb{Z}_6$.
$\therefore (\mathbb{Z}_6, + \mod 6, \cdot \mod 6)$ is not a field.

Remark 1.2.7. Proposition 1.2.2 is not a sufficient condition to conclude that an algebraic structure is a field. Consider the following example:

Example 1.2.6. Consider $(\mathbb{Z}, +, \cdot)$. Clearly $\forall \alpha, \beta \in \mathbb{Z}$ we have $\alpha\beta = 0 \iff \alpha = 0$ or $\beta = 0$. But $(\mathbb{Z}, +, \cdot)$ is not a field because there is no multiplicative inverse for $2 \in \mathbb{Z}$.

Let us consider some more nontrivial examples of algebraic structures and see which of the three axioms mentioned in remark 1.2.5 are satisfied.

Example 1.2.7. Let $\mathcal{F}(\mathbb{R}) = \{f \mid f : \mathbb{R} \to \mathbb{R}\}$, the set of all real-valued functions. Then $(\mathcal{F}(\mathbb{R}), +, \cdot)$ where $+$ and $\cdot$ are pointwise addition and multiplication of functions:

Given $f,\ g \in \mathcal{F}(\mathbb{R})$,
$+((f,\ g)) = (f+g)(x) = f(x) + g(x)$ and $\cdot((f,\ g)) = (f \cdot g)(x) = f(x) \cdot g(x)$.

Is $(\mathcal{F}(\mathbb{R}), +, \cdot)$ a field? If not, which axioms fail?

Answer: Since the binary operations defined on our set $F(\mathbb{R})$ are the point-wise addition and multiplication and the images are reals; commutativity , associativity and distributivity are preserved. Note $1_{\mathcal{F}(\mathbb{R})} = \mathbf{1}$ and $0_{F(\mathbb{R})} = \mathbf{0}$ where $\mathbf{1}$ and $\mathbf{0}$ are the constant function $I(x) = 1$ and zero function $Z(x) = 0$ respectively! However, multiplicative inverses may not exist for all elements.
Consider the function $f(x) = x$. Then $\left(f(x)\right)^{-1}$ does not exist. Suppose there is a $g(x) \in \mathcal{F}(\mathbb{R})$ such that $g(x)f(x) = \mathbf{1}$. Then the only candidate for $g(x)$ would be the functions of the form $G_\alpha = \begin{cases} \frac{1}{x} & \text{if } x \neq 0 \\ \alpha & \text{if } x = 0 \end{cases}$, where $\alpha \in \mathbb{R}$. So, for every $\alpha \in \mathbb{R}$ we can obtain a different function in $F(\mathbb{R})$! However, $G_\alpha(0)f(0) = \alpha.0 = 0$
Furthermore, the observation above should help us characterize all the nonzero elements in $F(\mathbb{R})$ that have multiplicative inverses. The functions that do not have a **zero** or **root**. The multiplicative inverse of such function $f(x)$ would simply be $\frac{1}{f(x)}$

Now given an algebraic structure, $(\mathcal{R}, +, \cdot)$ where the multiplicative commutativity may not be satisfied. Are there elements $x \in \mathcal{R}$ such that it commutes with every element in $\mathcal{R}$? This gives rise to the notion of the "center" of an algebraic structure.

Definition 1.2.4 (Center)**.** Given an algebraic structure $(\mathcal{R}, +, \cdot)$, we define the **center** of $\mathcal{R}$, $C(\mathcal{R})$, to be the set $\{x \in \mathcal{R} \mid xr = rx \ \forall r \in \mathcal{R}\}$

Remark 1.2.8. The $C(\mathcal{R})$ is always non-empty regardless of the structure $(\mathcal{R}, +, \cdot)$ since the additive identity is always in $C(\mathcal{R})$

Remark 1.2.9. The center of R possesses important algebraic properties; in particular it satisfies **ALL** the properties that $(\mathcal{R}, +, \cdot)$ satisfies. Hence, you can think of $(C(\mathcal{R}), +, \cdot)$ as an **algebraic substructure.**

We proceed to define such notion precisely.

Definition 1.2.5 (Substructure)**.** Given an algebraic structure $(\mathcal{R}, +, \cdot)$ an **algebraic substructure** $(\mathcal{S}, + \restriction \mathcal{S} \times \mathcal{S}, \cdot \restriction \mathcal{S} \times \mathcal{S})$ if $\mathcal{S} \subseteq \mathcal{R}$ and

1. The structure satisfies all the properties that the two tuple $(\mathcal{R}, +)$ satisfies including closure under $+$.
2. It is closed under the binary operation $\cdot$

NOTATION $+ \restriction \mathcal{S} \times \mathcal{S} \to \mathcal{S}$ and $\cdot \restriction \mathcal{S} \times \mathcal{S} \to \mathcal{S}$ are just restrictions maps. These notations basically indicate that if $x,\ y \in \mathcal{S}$, then $+((x,\ y)) = x + y \in \mathcal{S}$ and $\cdot((x,\ y)) = x \cdot y \in \mathcal{S}$.

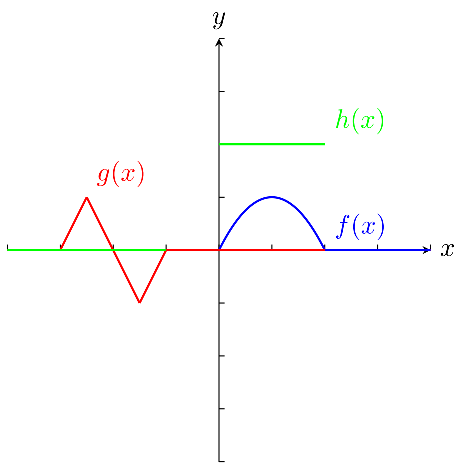
Remark 1.2.10. Note that closure under the binary operation $\cdot$ automatically implies that the multiplicative associativity and distributivity over $+$ are also satisfied since it is inherited from $\mathcal{R}$.

Remark 1.2.11. Now if an algebraic structure $(\mathcal{R}, +, \cdot)$ has a multiplicative identity $1_{\mathcal{R}}$ and the triple $(\mathcal{S}, + \restriction \mathcal{S} \times \mathcal{S}, \cdot \restriction \mathcal{S} \times \mathcal{S})$ is an algebraic substructure, this substructure does NOT need to possess the $1_{\mathcal{R}}$. Whereas $C(\mathcal{R})$ which is also a substructure of $\mathcal{R}$, will CONTAIN $1_{\mathcal{R}}$ by its definition. The following example will capture this important remark.

Exercise 1.2.4. Let $(\mathcal{R}, +, \cdot)$ be an algebraic structure and $\alpha \in \mathcal{R}$ has a multiplicative inverse and $\alpha \in C(\mathcal{R})$. Is $\alpha^{-1} \in C(\mathcal{R})$?

Example 1.2.8. Given a function $f : \mathbb{R} \to \mathbb{R}$ we say that f has a *compact support* if there exists $a < b$ such that $f(x) = 0\ \forall x \notin [a,\ b]$.
Consider the set $\text{supp}(\mathbb{R}) = \{f \in \mathcal{F}(\mathbb{R}) \mid f \text{ has compact support}\}$ with usual pointwise addition and pointwise multiplication. Is $\text{supp}(\mathbb{R})$ an algebraic substructure of $\mathcal{F}(\mathbb{R})$? What can you conclude about it?

Answer: First let us see some of the elements in the set.



Now let us determine whether $\text{supp}(\mathbb{R})$ is an algebraic substructure of $\mathcal{F}(\mathbb{R})$. Clearly the identity element is simply $z(x) = 0$. $z(x)$ has a compact support since if we pick $a = 0$ and $b = 1$ then $z(x) = 0\ \forall x \notin [0,\ 1]$. Since addition and multiplication are pointwise and take place in $\mathbb{R}$, commutativity, associativity and distributivity hold. We only need to show the closure of $\text{supp}(\mathbb{R})$ under these operations! That is if we add or multiply two functions $K(x)$, and $G(x)$ in $\text{supp}(\mathbb{R})$, the result will still be in $\text{supp}(\mathbb{R})$.

Let $K(x),\ G(x) \in \text{supp}(\mathbb{R}) \implies \exists\ a < b$ and $a' < b'$ in $\mathbb{R}$ such that $K(x) = 0\ \forall x \notin [a,\ b]$ and $G(x) = 0\ \forall x \notin [a',\ b']$. Let $M = \max\{b,\ b'\}$ and $N = \min\{a, a'\}$. Then it can easily be verified that $(K + G)(x) = K(x) + G(x)$ has $[N,\ M]$ as its compact support. Since if $x_0 > M \geq b$ and $M \geq b'$ then we have that $K(x_0) = 0$ and $G(x_0) = 0$ hence $(K + G)(x_0) = 0$. If $x_1 < N \leq a'$ and $N \leq a$ again we have that $K(x_1) = 0 = G(x_1)$. Hence, $(G + K)(x_1) = 0$.
$\therefore (G + K)(x) \in \text{supp}(\mathbb{R})$. Similar arguments can be made to show that $(K \cdot G)(x) \in \text{supp}(\mathbb{R})$.
$\therefore$ By definition 1.2.5, $(\text{supp}(\mathbb{R}),\ +,\ \cdot)$ is an algebraic substructure of $(\mathcal{F}(\mathbb{R}),\ +,\ \cdot)$.

Remark 1.2.12. We have already concluded that $1_{\mathcal{F}(\mathbb{R})} = I(x) = 1$ and $\mathcal{F}(\mathbb{R})$ does not satisfy the existence of multiplicative inverse axiom, even though there are functions that have multiplicative inverses; namely any functions that does not have a root.
So, now what is $1_{\text{supp}(\mathbb{R})}$? Is it still $I(x) = 1$?
It's obvious that $I(x) \cdot f(x)$ for any $f(x) \in \text{supp}(\mathbb{R})$ will be equal to $f(x)$, but is $I(x) = 1 \in \text{supp}(\mathbb{R})$? The answer is no, that's because such constant function does not have any compact support. As illustrated by our graphs- namely by $f(x)$, $h(x)$- it is true for every function there is a function whose product will be the original function. But there is no function that works for ALL the functions.

Theorem 1.2.2. *The triple $(\mathbb{Z}_p, + \mod p, \cdot \mod p)$ is a field* iff *p is a prime.*

Exercise 1.2.5. Prove **Theorem 1.2.2**. Use the fact that if two integers m, n are relatively prime, there exists integers r and s such that $mr + ns = 1$.

Example 1.2.9. Does the set of functions with infinitely many roots form an algebraic substructure of $\mathcal{F}(\mathbb{R})$?
We leave the answer as an exercise. Hint $f(x) = \begin{cases} 1 \text{ if } x \in \mathbb{Q} \\ 0 \text{ if } x \in \mathbb{I} \end{cases}$ These type of functions are modifications of

Dirichlet Functions

Exercise 1.2.6. Prove the statement in the above example.

Definition 1.2.6 (Polynomial Structure)**.** Let $(\mathcal{R}, +, \cdot)$ be any algebraic structure.
Then $\mathcal{R}[x] = \left\{ \sum_{i=0}^{n} \alpha_i x^i \mid n \in \mathbb{Z}_{\geq 0} \text{ and } \forall\ 0 \leq i \leq n, \alpha_i \in \mathcal{R} \right\}$ is called a **polynomial structure** with the following operations:
Let $p(x) = \sum_{i=0}^{n} \alpha_i x^i$, $q(x) = \sum_{i=0}^{m} \beta_i x^i, n \leq m$.

$\boxplus : \mathcal{R}[x] \times \mathcal{R}[x] \longrightarrow \mathcal{R}[x]$

$$(p(x),\ q(x)) \longmapsto p \boxplus q = \sum_{i=0}^{n} \underbrace{(\alpha_i + \beta_i)}_{\text{Note: } + \text{ is the binary operation defined } \mathcal{R}} x^i + \sum_{i=n+1}^{m} \beta_i x^i$$

$\star : \mathcal{R}[x] \times \mathcal{R}[x] \longrightarrow \mathcal{R}[x]$

$$\begin{aligned}(p(x),\ q(x)) \longmapsto p \star q &= (\alpha_0 + \alpha_1 x + \cdots + \alpha_n x^n) \underbrace{\cdot}_{\text{multiplication for } \mathcal{R}} (\beta_0 + \beta_1 x + \cdots + \beta_m x^m) \\ &= \alpha_0\beta_0 + \alpha_0\beta_1 x + \cdots + \alpha_0\beta_m x^m \\ &\quad + \alpha_1\beta_0 x + \alpha_1\beta_1 x^2 + \cdots + \alpha_1\beta_m x^{m+1} \\ &\quad \vdots \\ &\quad + \alpha_n\beta_0 x^n + \alpha_n\beta_1 x^{n+1} + \cdots + \alpha_n\beta_m x^{n+m} \\ &= \sum_{i=0}^{n} \left(\sum_{j=0}^{m} \alpha_i\beta_j x^{i+j} \right)\end{aligned}$$

Remark 1.2.13. Note when we distributed the terms and expanded everything in the product $p(x) \star q(x)$ technically we should have obtained a different format. For instance, the first occurrences in the distribution which are of significance would be $\alpha_1 x \beta_0$ and $\alpha_1 x \beta_1 x$. But since such a form does not satisfy the format of a polynomial, we defined it to be $\alpha_1\beta_0 x$ and $\alpha_1\beta_1 x^2$ for things to make more sense. But note that, $\alpha_1\beta_1 x^2 \neq \beta_1\alpha_1 x^2$ since the underlying set $\mathcal{R}$ may not be commutative. (*ie*, $\cdot$ may not be commutative)

Remark 1.2.14. You can easily verify that the triple $(\mathcal{R}[x], \boxplus, \star)$ forms an special "algebraic structure." Special because of the previous remark. We swapped β_1 and x to go from $\alpha_1 x \beta_1 x$ to $\alpha_1\beta_1 x^2$, but we are not swapping α_1 and β_1. Hence, this idea of a **POLYNOMIAL STRUCTURE** is really significant and people talk about it *iff* the underlying set is commutative with unity.

Example 1.2.10. Consider the algebraic structure $(\mathcal{M}_{2\times 2}(\mathbb{R}), +, \cdot)$, where $+$ and $\cdot$ are the usual matrix addition and multiplication.

Compute $p(x) \boxplus q(x)$ and $p(x) \star q(x)$ in $(\mathcal{M}_{2\times 2}(\mathbb{R})[x], \boxplus, \star)$, where $p(x) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 2 & 1 \\ 1 & 0 \end{bmatrix} x$, and $q(x) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} x^3 + \begin{bmatrix} 1 & 0 \\ \frac{1}{2} & \sqrt{2} \end{bmatrix} x$. Does this algebraic structure have a multiplicative identity? Does it have additive identity?

Answer: $p(x) \boxplus q(x) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 3 & 1 \\ \frac{3}{2} & \sqrt{2} \end{bmatrix} x + \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} x^3$

$p(x) \star q(x) = \begin{bmatrix} 1 & 0 \\ \frac{1}{2} & \sqrt{2} \end{bmatrix} x + \begin{bmatrix} \frac{5}{2} & \sqrt{2} \\ 1 & 0 \end{bmatrix} x^2 + \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} x^3 + \begin{bmatrix} 4 & 2 \\ 2 & 0 \end{bmatrix} x^4$

The constant polynomial $I(x) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ is the **multiplicative identity** and the constant polynomial $z(x) = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$ is the **additive identity**, or also called the **zero element**.

Exercise 1.2.7. Is $\star$ commutative in example 1.2.10? Which polynomials have multiplicative inverses?

Example 1.2.11. Consider the field $(\mathbb{R}, +, \cdot)$, (which is an algebraic structure) Then consider its corresponding polynomial structure $(\mathbb{R}[x], \boxplus, \star)$. So, $\mathbb{R}[x]$ is basically the set of ALL possible polynomials whose coefficients are reals. Then the binary operations $\boxplus$ and $\star$ are the usual addition and multiplication of polynomials. The only polynomials that have multiplicative inverses are the nonzero constant polynomials.

Remark 1.2.15. If $\mathcal{R}$ has a multiplicative identity so does $\mathcal{R}[x]$. Furthermore, if $\alpha \in \mathcal{R}$ has multiplicative inverse, then so does the constant polynomial $p(x) = \alpha x^0 = \alpha_0$, in particular $p^{-1}(x) = \alpha_0^{-1} x^0 = \alpha_0^{-1}$. Do not forget α^{-1} denotes the multiplicative inverse. It is not the usual $\frac{1}{\alpha}$. That may not even makes sense. Think α could be a matrix.

2 Vector Spaces

2.1 Basics

Definition 2.1.1 (Vector Space)**.** Let $(\mathbb{F}, +, \cdot)$ be a field and let V be a nonempty set equipped with the following two binary operations:

1. Vector Addition

$$\begin{aligned} \boxplus : V \times V &\to V \\ (x, y) &\longmapsto x \boxplus y \end{aligned}$$

2. Scalar Multiplication

$$\begin{aligned} \odot : \mathbb{F} \times V &\to V \\ (c, v) &\longmapsto c \odot v \end{aligned}$$

We say the quadruple $\mathbb{V} = (V, \boxplus, \odot, (\mathbb{F}, +, \cdot))$ is a **vector space over** $\mathbb{F}$ if the following properties are satisfied:

1. $\boxplus$ is commutative
2. $\boxplus$ is associative
3. **additive identity**: There is a vector in V, denoted by $\underline{0}$ s.t. for all $v \in V$ we have that

$$v \boxplus \underline{0} = v$$

4. **additive inverse**: For all $v \in V$ there is a vector $-v \in V$ such that

$$v \boxplus (-v) = \underline{0}$$

$\underline{0}$ is called the **zero vector** of V

5. $\odot$ is **associative over scalar multiplication.** That is for all $\alpha, \beta \in \mathbb{F}$ and $v \in V$ we have

$$\underbrace{\alpha \cdot \beta}_{\text{multiplication in } \mathbb{F}} \qquad (\alpha.\beta) \odot v = \alpha \odot (\beta \odot v)$$

6. **distributivity of $\odot$ over scalar addition**: For all $\alpha, \beta \in \mathbb{F}$ and $v \in V$

$$(\alpha + \beta) \odot v = (\alpha \odot v) \boxplus (\beta \odot v)$$

7. **distributivity of $\odot$ over vector addition**: For all $\alpha \in \mathbb{F}$ and $v, w \in V$, we have

$$\alpha \odot (v \boxplus w) = (\alpha \odot v) \boxplus (\alpha \odot w)$$

8. For all $v \in V$ we have

$$1_{\mathbb{F}} \odot v = v$$

Furthermore, we call an element $v \in V$ a **vector** and $\alpha \in \mathbb{F}$ a **scalar**.

Theorem 2.1.1. *Let $(\mathbb{F}, +, \cdot)$ be a field. Then there is a vector space over $\mathbb{F}$*

Proof. Define $V := \mathbb{F}$ and $\boxplus := +$ and $\odot := \cdot$. It should be very easy to verify why all the axioms of vector spaces would be satisfied now. ■

Example 2.1.1. Consider the fields $(\mathbb{Q}, +, \cdot)$ and $(\mathbb{R}, +, \cdot)$ where $+$ and $\cdot$ are the usual addition and multiplication. Is $(\mathbb{Q}, +, \cdot, (\mathbb{R}, +, \cdot))$ a vector space over $\mathbb{R}$?

Answer: No. $\mathbb{Q}$ will not be closed under the binary operation $\cdot$. Note that $\cdot((\sqrt{2}, 2)) = 2\sqrt{2} \notin \mathbb{Q}$ However, $(\mathbb{R}, +, \cdot, (\mathbb{Q}, +, \cdot))$ is a vector space over $\mathbb{Q}$.

Example 2.1.2. Let $(\mathbb{F}, +, \cdot)$ be any field.

Take the set of all $m \times n$ matrices:

$$\mathcal{M}_{m\times n}(\mathbb{F}) = \left\{ \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & & & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} : a_{ij} \in \mathbb{F} \text{ for all } 1 \leq i \leq m, 1 \leq j \leq n \right\}$$

Again to define vector addition and scalar multiplication on $M_{m\times n}$ we take the 'natural' operation, that is

$$\hat{A} \boxplus \hat{B} = \underbrace{[a_{ij} + b_{ij}]_{m\times n}}_{\text{NOTE: } + \text{ is the addition in } \mathbb{F}}$$

Where $\hat{A} = [a_{ij}]_{m\times n}$ and $\hat{B} = [b_{ij}]_{m\times n}$ and

$$\alpha \odot \hat{A} = \underbrace{[\alpha \cdot a_{ij}]_{m\times n}}_{\text{NOTE: } \cdot \text{ is the multiplication in } \mathbb{F}}$$

where $\alpha \in \mathbb{F}$.

Example 2.1.3. Consider the field $(\mathbb{Z}_5, + \mod 5, \cdot \mod 5)$. Now $(\mathcal{M}_{3\times 3}(\mathbb{Z}_5), \boxplus, \odot, (\mathbb{Z}_5, + \mod 5, \cdot \mod 5))$ forms a vector space equipped with the vector addition and scalar multiplication as defined as in example 2.1.2. Determine $\begin{bmatrix} 0 & 1 & 1 \\ 2 & 2 & 4 \\ 3 & 1 & 2 \end{bmatrix} \boxplus \begin{bmatrix} 1 & 2 & 1 \\ 2 & 3 & 4 \\ 2 & 3 & 2 \end{bmatrix}$ and $3 \odot \begin{bmatrix} 1 & 2 & 1 \\ 2 & 3 & 4 \\ 2 & 3 & 2 \end{bmatrix}$. In addition, what is the additive inverse of $\begin{bmatrix} 1 & 2 & 1 \\ 2 & 3 & 4 \\ 2 & 3 & 2 \end{bmatrix}$?

Answer: $\begin{bmatrix} 0 & 1 & 1 \\ 2 & 2 & 4 \\ 3 & 1 & 2 \end{bmatrix} \boxplus \begin{bmatrix} 1 & 2 & 1 \\ 2 & 3 & 4 \\ 2 & 3 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 3 & 2 \\ 4 & 0 & 3 \\ 0 & 4 & 4 \end{bmatrix}$ and $3 \odot \begin{bmatrix} 1 & 2 & 1 \\ 2 & 3 & 4 \\ 2 & 3 & 2 \end{bmatrix} = \begin{bmatrix} 3 & 1 & 3 \\ 1 & 4 & 2 \\ 1 & 4 & 1 \end{bmatrix}$ Lastly, the additive inverse of $\begin{bmatrix} 1 & 2 & 1 \\ 2 & 3 & 4 \\ 2 & 3 & 2 \end{bmatrix}$ is $-\begin{bmatrix} 1 & 2 & 1 \\ 2 & 3 & 4 \\ 2 & 3 & 2 \end{bmatrix} = \begin{bmatrix} 4 & 3 & 4 \\ 3 & 2 & 1 \\ 3 & 2 & 3 \end{bmatrix}$

Exercise 2.1.1. If we defined the vector addition to be $\hat{A} \boxplus \hat{B} = [a_{ij}.b_{ij}]$ in example 2.1.2, would $(\mathcal{M}_{m\times n}(\mathbb{F}), \boxplus, \odot, (\mathbb{F}, +, \cdot))$ still be a vector space over $\mathbb{F}$?

Example 2.1.4. Another example of a vector space is related to the set $\mathcal{F}(\mathbb{R})$. Let $V = \mathcal{F}(\mathbb{R}) = \{f | f : \mathbb{R} \to \mathbb{R}\}$ and the field $\mathbb{F} = \mathbb{R}$ with the usual addition and multiplication. Now given any two functions $f, g \in \mathcal{F}(\mathbb{R})$ and $c \in \mathbb{R}$, we define vector addition as

$$\boxplus\big(f(x), g(x)\big) = f \boxplus g = (f+g)(x) = f(x) + g(x)$$

So again it is simply component-wise addition in $\mathbb{R}$.

Define scalar multiplication as the following:

$$c \odot f = (cf)(x) = cf(x)$$

So scalar multiplication is simply a vertical stretch by a factor of c.

Example 2.1.5. Let $\mathbb{Z}_7[x] = \left\{ \sum_{i=0}^{n} a_i x^i \,\middle|\, a_i \in \mathbb{Z}_7 \text{ and } n \in \mathbb{Z}_{\geq 0} \right\}$, which represents the set of all polynomials whose coefficient are from $\mathbb{Z}_7$. Let $\mathbb{V} = (\mathbb{Z}_7[x], \boxplus, \odot, (\mathbb{Z}_7, + \mod 7, \cdot \mod 7))$. Define the binary operations $\boxplus$ and $\odot$ so the above quadruple constitutes a vector space over $\mathbb{Z}_7$.

Answer: Let $p(x) = \sum_{i=0}^{n} \alpha_i x^i$, $q(x) = \sum_{i=0}^{m} \beta_i x^i \in \mathbb{Z}_7[x], n \leq m$. Let $\gamma \in \mathbb{Z}_7$.
Define $\boxplus : \mathbb{Z}_7[x] \times \mathbb{Z}_7[x] \longrightarrow \mathbb{Z}_7[x]$

$$(p(x),\ q(x)) \longmapsto p(x) \boxplus q(x) = \sum_{i=0}^{n} ((\alpha_i + \beta_i) \mod 7)x^i + \sum_{i=n+1}^{m} \beta_i x^i$$

$\odot : \mathbb{Z}_7 \times \mathbb{Z}_7[x] \longrightarrow \mathbb{Z}_7[x]$

$$(\gamma,\ p(x)) \longmapsto \gamma \odot p(x) = \sum_{i=0}^{n} ((\gamma\alpha_i) \mod 7)x^i$$

For instance, if $p(x) = 6x^7 + 4x^3 + 2x + 4$ and $q(x) = 5x^4 + 3x^3 + 2x^2 + 4x + 5$,
then $p(x) \boxplus q(x) = 6x^7 + 5x^4 + 2x^2 + 6x + 2$
Also, $-p(x) = x^7 + 3x^3 + 5x + 3$ and $4 \odot p(x) = 3x^7 + 2x^3 + x + 2$

Theorem 2.1.2. *Let $\mathbb{V} = (V, \boxplus, \odot, (\mathbb{F}, +, \cdot))$ be a vector space over field $\mathbb{F}$. The following properties hold:*

1. *The additive identity of V is unique*
2. *The additive inverse of any $v \in V$ is unique*
3. *For all $v \in V$,*
$$0 \odot v = \underline{0}$$
Where 0 is the additive identity of the field $\mathbb{F}$ and $\underline{0}$ is the additive identity of V
4. *For all scalars $\alpha \in \mathbb{F}$,*
$$\alpha \odot \underline{0} = \underline{0}$$
5. *For all $v \in V$ and $c \in \mathbb{F}$,*
$$(-c) \odot v = c \odot (-v) = -1 \odot (c \odot v)$$

Proof. We proceed to prove 1. and 3.

- Let $\underline{0}$ and $\underline{0}'$ be two additive zeros in V. Then we have that for all $v \in V$,
$$v \boxplus \underline{0} = v \tag{1}$$
$$v \boxplus \underline{0}' = v \tag{2}$$
Start off with $\underline{0}$. Then by (1) one of the vectors $v \in V$ is $\underline{0}'$. So we get that
$$\underline{0}' \boxplus \underline{0} = \underline{0}'$$
Now by (2) $\underline{0}$ is also an element of V and $\underline{0}'$ is also an additive identity of V. So we also get
$$\underline{0} \boxplus \underline{0}' = \underline{0}$$
Since we know $\boxplus$ is commutative, this means
$$\underline{0}' = \underline{0}' \boxplus \underline{0} = \underline{0} \boxplus \underline{0}' = \underline{0}$$

- Take $0 \in \mathbb{R}$ and $v \in V$. We know that $0 + 0 = 0$, which allows us to see that
$$0 \odot v = (0 + 0) \odot v$$
Now by distributivity over scalar addition, we have
$$0 \odot v = (0 + 0) \odot v = (0 \odot v) \boxplus (0 \odot v) \tag{3}$$
We know that V is closed under $\odot$, hence $0 \odot v$ is some element of V. By the properties of vector spaces we know that every element has an additive inverse. Denote $-(0 \odot v)$ as its additive inverse. Then applying this to both sides of the equation (3), we get the following result:
$$0 \odot v \boxplus -(0 \odot v) = (0 \odot v \boxplus 0 \odot v) \boxplus \big(-(0 \odot v)\big)$$
$$\begin{aligned}\implies \underline{0} &= 0 \odot v \boxplus \big(0 \odot v \boxplus (-(0 \odot v))\big) && \text{By associativity of } \boxplus \\ &= 0 \odot v \boxplus \underline{0} \\ &= 0 \odot v && \text{By additive identity}\end{aligned}$$

■

Exercise 2.1.2. Prove properties 2., 4. and 5. of **Theorem 2.1.2**

The following 3 subsections will be a quick review of MATA22 materials. You are expected to know all the theorems/definitions etc.

2.2 Subspaces

Definition 2.2.1 (Subspace)**.** A subset W of the vector space $(V, \boxplus, \odot, (\mathbb{F}, +, \cdot))$ over a field $\mathbb{F}$, is a **subspace** of V if W itself is a vector space. *i.e*, it fulfills the requirements of a vector space.

Given a subset W of our vector space V, determining whether it is also a vector space (and hence a subspace) of V might be tedious or cumbersome since it would require us to go over all eight axioms associated with the corresponding or inherited vector addition and scalar multiplication. However, the following 'theorem' facilitates a simple test for determining whether or not W is a subspace of V:

Theorem 2.2.1 (Subspace Test)**.** *Let W be a non-empty subset of a vector space V. Then W is a subspace of V if and only if it is closed under vector addition and scalar multiplication. That is, the following holds:*

1. *for all $x, y \in W$,*
$$x \boxplus y \in W$$
2. *for all $\alpha \in \mathbb{F}$ and $x \in W$,*
$$\alpha \odot x \in W$$

To see a proof note that for instance condition 2 of theorem 2.2.1 and statement 3 of theorem 2.1.2 imply that additive identity is an element of W. So, we have that $0 \odot v = \underline{0} \in W$

Now given $v \in W$, how about the additive inverse of v? Here, we need statement 5 of theorem 2.1.2, along with closure of W under $\odot$.

Let us now construct some subspaces from a given vector space $(V, \boxplus, \odot, (\mathbb{F}, +, \cdot))$. Consider the following 4 **Theorems.**

Definition 2.2.2 (Sum of Vector Spaces)**.** Let V be a vector space. Let $A, B \subseteq V$ be any two *<u>subsets</u>* of V. We define the sum of these two sets (with respect to the vector addition in V) as the following:

$$A + B := \{a \boxplus b : a \in A, b \in B\}$$

Theorem 2.2.2. *Suppose W_1 and W_2 are subspaces of a vector space V. Then $W_1 + W_2$ is also a subspace.*

Exercise 2.2.1. Prove **Theorem 2.2.2**

Definition 2.2.3 (Span)**. span** of S is the set of ALL possible linear combinations of elements in S. That is, if $S = \{s_1, s_2, \ldots, s_n\}$, then $\text{sp}(S) = \{\sum_{i=1}^{n} \alpha_i s_i | \, \alpha_i \in \mathbb{F}, s_i \in S \text{ for all } 1 \leq i \leq n\}$

Remark 2.2.1. Let V be a vector space. If $S = \emptyset$ then $sp(S) = \{\underline{0}\}$

Remark 2.2.2. S does not need to be finite. However, even if S is infinite we only consider all finite possible linear combinations of elements of S. It does not make sense to talk about an infinite sum of vectors for our purposes.

Theorem 2.2.3. *If $S \subseteq V$ then $\text{sp}(S) \subseteq V$ is a subspace of V.*

Exercise 2.2.2. Prove **Theorem 2.2.3**

Theorem 2.2.4. *Let J be any index set.(It is a set whose elements are used to enumerate the elements of set.) and let $\{W_\alpha : \ \alpha \in J\}$ be a collection of subspaces of V. Then $\bigcap\limits_{\alpha \in J} W_\alpha$ is a subspace of V.*

Exercise 2.2.3. Prove **Theorem 2.2.4**

Remark 2.2.3. How about if replace the intersection with the union? Note that then $\bigcup\limits_{\alpha \in J} W_\alpha$? Then answer is no! For instance consider $\mathbb{R}[x] = \left\{ \sum_{i=0}^{n} \alpha_i x^i : \ \forall \ 0 \leq i \leq n \ \alpha_i \in \mathbb{R} \ , n \in \mathbb{Z}_{\geq 0} \right\}$ the set of all polynomials. Define the vector addition to be the pointwise addition of coefficients associated to the variable x with the same power, and scalar multiplication over the field $\mathbb{R}$ to be the usual multiplication distributed to all the coefficients. Then the set of all **even** polynomials $W_1 = \left\{ \sum_{i=0}^{n} \alpha_i x^{2i} : \ \forall \ 0 \leq i \leq n \ \alpha_i \in \mathbb{R} \ , n \in \mathbb{Z}_{\geq 0} \right\}$ and the set of all **odd** polynomials $W_2 = \left\{ \sum_{i=0}^{n} \alpha_i x^{2i+1} : \ \forall \ 0 \leq i \leq n \ \alpha_i \in \mathbb{R} \ , n \in \mathbb{Z}_{\geq 0} \right\}$ are two subspaces of $\mathbb{R}[x]$. Note that the zero vector $z(x) = 0$ is both an even and odd polynomial. In addition, the zero vector, $z(x) = 0$, is the only vector common in both. However, $W_1 \cup W_2$ is not a subspace. It is not closed under vector addition. Take $p(x) = x^2 + 1 \in W_1$ and $q(x) = x^3 + 2x \in W_2$. However, $p(x) + q(x) = x^3 + 2x + x^2 + 1$ is neither **even** nor **odd** polynomial, Hence, $W_1 \cup W_2$ is not closed under vector addition.

If we impose the following condition then the union of subspaces of a given vector space will be a subspace.

Theorem 2.2.5. *Let $W_1 \subseteq W_2 \subseteq W_3 \subseteq \cdots$ be an increasing chain of subspaces of V. Then $\bigcup\limits_{n \in \mathbb{N}} W_n$ is a subspace.*

Exercise 2.2.4. Prove **Theorem 2.2.5**

Now the following notion will allow us to study the structure of a given vector space V from within the vector space. From within we will find subspaces that will be "extracted" in such a way that will be **isomorphic** to V.

We will study it in more detail once we talk about linear transformation and study the notion of isomorphism. First let us see just a definition of it.

Definition 2.2.4 (Direct Sum)**.** Let V be a vector space over a field $\mathbb{F}$. Suppose $W_1, \ldots, W_k$ are subspaces of V such that for all $1 \leq i \neq j \leq k$,

$$W_i \cap \big(\sum_{j \neq i} W_j\big) = \{\underline{0}\}$$

(that means no non-zero vector in W_i can be expressed as sum of vectors from other subspaces W_j) We say V is a **direct sum** of $W_1, \ldots, W_k$ if

$$V = W_1 + W_2 + \cdots + W_k$$

That is to say if every vector v in V can be expressed uniquely as

$$v = w_1 + w_2 + \cdots + w_k$$

where $w_i \in W_i$ for all $1 \leq i \leq k$. We then denote V by $V = W_1 \bigoplus W_2 \bigoplus \cdots \bigoplus W_k = \bigoplus_{i=1}^{k} W_i$.

Example 2.2.1. Let $V = \mathbb{R}^5$. Let

$$\begin{aligned} W_1 &= \{(\alpha, 0, 0, 0, 0) : \alpha \in \mathbb{R}\} \\ W_2 &= \{(0, \alpha, \beta, 0, 0) : \alpha, \beta \in \mathbb{R}\} \\ W_3 &= \{(0, 0, 0, \alpha, \beta) : \alpha, \beta \in \mathbb{R}\} \end{aligned}$$

Then $V = W_1 \bigoplus W_2 \bigoplus W_3$.
Proof: It's clear that for any i , $1 \leq i \leq 3$

$$W_i \cap \big(\sum_{j \neq i} W_j\big) = \{(0, 0, 0, 0, 0) = \underline{0}\}$$

It can also be easily checked that the subspaces are pairwise disjoint as well. And given any vector $v \in \mathbb{R}^5$, $v = (v_1, v_2, \ldots, v_5)$ where $v_i \in \mathbb{R}$ for all $1 \leq i \leq 5$, we can see that

$$v = \underbrace{(v_1, 0, 0, 0, 0)}_{\in W_1} + \underbrace{(0, v_2, v_3, 0, 0)}_{\in W_2} + \underbrace{(0, 0, 0, v_4, v_5)}_{\in W_3}$$

2.3 Bases

Definition 2.3.1 (Linear Dependence)**.** Let V be a vector space over a field $\mathbb{F}$. A set of vectors $S = \{v_1, \dots v_n\} \subseteq V$ is **linearly dependent** if there are some non-zero scalars, α_i, such that

$$\sum_{i=1}^{n} \alpha_i v_i = \underline{0}$$

If there are no such non-zero scalars, i.e, the only way to write the sum above is when all $\alpha_i = 0$ for $1 \leq i \leq n$, we say that the set of vectors in S are **linearly independent**. Hence, S is linearly independent if whenever $\alpha_i \in \mathbb{F}$ and $v_i \in S$ such that $\sum_{i=1}^{n} \alpha_i v_i = \underline{0}$, then $\alpha_i = 0$ for all i.

Remark 2.3.1. If a set of vectors S is linearly dependent, then there is a vector $v_j \in S$ that can be written as a linear combination of some other vectors in S. That is because there is at least one non-zero scalar $\alpha_j \in \mathbb{F}$ such that

$$\sum_{i=1}^{n} \alpha_i v_i = \underline{0} \tag{1}$$

Then adding the additive inverse of each vector $\alpha_i v_i$, for all i except for $i = j$ to both sides of equation (1) we get that

$$\alpha_j v_j = -\alpha_1 v_1 - \cdots - \alpha_{j-1} v_{j-1} - \alpha_{j+1} v_{j+1} - \cdots - \alpha_n v_n$$

Now since α_j is a non-zero scalar it has a multiplication inverse. Hence, we have that

$$v_j = \alpha_j^{-1}\big(- \alpha_1 v_1 - \cdots - \alpha_{j-1} v_{j-1} - \alpha_{j+1} v_{j+1} - \cdots - \alpha_n v_n\big)$$

Remark 2.3.2. S does not have to be finite in order for it to be linearly independent/dependent. It depends on the vector space V.

Exercise 2.3.1. Show that the set $\{e^x, e^{2x}, e^{3x}\}$ is linearly independent in the vector space of $\mathcal{F}(\mathbb{R})$ over $\mathbb{R}$.

The notion of **basis** for a given vector space is extremely important. Not only can we generate the whole space using the vectors in this particular collection, but we can also use them to determine many algebraic properties as when we define certain maps from a vector space to another vector space.

Definition 2.3.2 (Basis)**.** Let V be a vector space over $\mathbb{F}$. A set of vectors $\mathcal{B}$ in V is a basis for V if the following conditions are met:

1. The set of vectors $\mathcal{B}$ spans V. We denote it by $\mathrm{sp}(\mathcal{B}) = V$
2. The set of vectors in $\mathcal{B}$ is linearly independent.

Remark 2.3.3. If $\{s_1, \dots, s_n\}$ is a spanning set for a vector space V and $\{v_1 \dots, v_m\}$ is linearly independent subset of V, then $m \leq n$.

Theorem 2.3.1. *Every vector space has a basis.*

Example 2.3.1. Consider the vector space $V = \mathbb{P}_n[x]$ over $\mathbb{R}$. Then what is a basis of V?

ANS: A basis for V is $\mathcal{B} = \{1, x, x^2, \ldots, x^n\}$. It's clear that $\text{sp}(\mathcal{B}) = \mathbb{P}_n[x]$ and the set of vectors is linearly independent, since we can't write any element in the form of others in the set through only sums and scalar multiplications.

Notation: $\mathbb{P}_n[x]$ is the set of all polynomials of degree at most n. $\mathbb{R}[x]$ represents the set of all polynomials of ALL possible degrees.

Theorem 2.3.2. *Let $\mathcal{B}$ be a set of non-zero vectors in a vector space V. Then $\mathcal{B}$ is a basis for V if and only if each vector v in V can be uniquely expressed as a linear combination of vectors in $\mathcal{B}$.*

Proof. First off, note that we require that $\underline{0} \notin \mathcal{B}$. This is because any subset of any vector space containing the $\underline{0}$ vector (or additive identity) is linearly dependent, since for any $r \in \mathbb{F}$, we have

$$r \cdot \underline{0} = \underline{0}$$

($\Rightarrow$)
Now assume $\mathcal{B} = \{b_1, \ldots, b_n\}$ is a finite basis for V. Choose a vector $v \in V$. Since β is a basis for V, we know that $\text{sp}(\mathcal{B}) = V$, so there exists some $\alpha_i \in \mathbb{F}$ such that

$$v = \sum_{i=1}^{n} \alpha_i b_i \quad (1)$$

Suppose v can also be expressed as another different linear combination of vectors in $\mathcal{B}$ for a contradiction. So we have that there exists $\gamma_i \in \mathbb{F}$ such that

$$v = \sum_{i=1}^{n} \gamma_i b_i \quad (2)$$

Subtracting (2) from (1) we get the following:

$$\sum_{i=1}^{n} (\alpha_i - \gamma_i) b_i = \underline{0}$$

Note here we have applied the additive inverse of each α_i, γ_i to both sides. Also know that $\alpha_i - \gamma_i \in \mathbb{F}$, so we get the relation

$$(\alpha_1 - \gamma_1) b_1 + (\alpha_2 - \gamma_2) b_2 + \cdots + (\alpha_n - \gamma_n) b_n = \underline{0}$$

But since $\mathcal{B}$ is a basis, we know that the elements in $\mathcal{B} = \{b_1, \ldots, b_n\}$ form a linearly independent set. So, we must have that $\alpha_i - \gamma_i = 0$ for all $1 \le i \le n$. I.e, we must have that $\alpha_i = \gamma_i$, which means (1) and (2) are the same. Since these were two arbitrary linear combinations, we thus get that v is uniquely expressed as a linear combination of vectors in $\mathcal{B}$.
($\Leftarrow$)
Now assume that for every vector $v \in V$ can be expressed uniquely as a linear combination of vectors in $\mathcal{B} = \{b_1, \ldots, b_n\}$. We proceed to show that $\mathcal{B}$ is a basis for V. We already have the first condition satisfied, since

every element in V is written as a linear combination of the vectors in $\mathcal{B}$, so they are all in $\mathrm{sp}(\mathcal{B})$, and $\mathcal{B}$ consists of vectors in V. So we get both directions of set containment and thus $\mathrm{sp}(\mathcal{B}) = V$. We need to show now that $\{b_1, \ldots, b_n\}$ is linearly independent. Let $\alpha_1, \ldots, \alpha_n \in \mathbb{F}$ such that

$$\alpha_1 b_1 + \alpha_2 b_2 + \cdots + \alpha_n b_n = \underline{0} \quad (3)$$

By our assumption every vector $v \in V$ can be expressed uniquely as a linear combination of vectors in $\mathcal{B}$. So this is also true for the zero vector, since it is also an element of V! Thus we must have that $\alpha_i = 0$ for all $1 \leq i \leq n$, for if there were one $\alpha_j \neq 0$ then (3) becomes

$$\alpha_j b_j = \underline{0} = 1 \cdot \underline{0}$$

Note that we cannot have $\alpha_j = 1$ since by the 8^{th} axiom of the vector spaces we have to have that $1 \cdot b_j = b_j$. Also, $\mathcal{B}$ only consists of non-zero vectors so, $b_j \neq 0$. Hence, $\underline{0}$ can be expressed as two different linear combinations of vectors in $\mathcal{B}$, which is a contradiction. So, we must have that $\alpha_i = 0$ for all i and hence, $\mathcal{B}$ is linearly independent. $\therefore$ $\mathcal{B}$ is a basis for V, as wanted. ■

Definition 2.3.3 (Dimension)**.** Let V be a vector space and $\mathcal{B}$ a basis for V. The number of elements in $\mathcal{B}$ is the **dimension** of V and is denoted $\dim(V)$.

Remark 2.3.4. The number of vectors in $\mathcal{B}$ is denoted by $|\mathcal{B}|$.

Definition 2.3.4 (Finitely Generated Vector Space)**.** If $|\mathcal{B}| < \infty$ (is finite) then V is said to be **finitely generated**.

2.4 Linear Transformations

Definition 2.4.1 (Linear Transformation)**.** Let V, W be two vector spaces over the same field $\mathbb{F}$. A function T that maps V to W is a **linear transformation** if it satisfies the following two criteria:

1. Preservation of addition: $T(u +_V v) = T(u) +_W T(v)$ for all $u, v \in V$
2. Preservation of scalar multiplication: $T(r \cdot_V u) = r \cdot_W T(u)$ for all $u \in V$ and $r \in \mathbb{F}$

Definition 2.4.2 (Domain & Co-Domain)**.** For a linear transformation $T : V \to W$, the vector space V is the **domain**, denoted by $\mathrm{dom}(T)$, and W is the **co-domain**.

Definition 2.4.3 (Image & Range)**.** Let $T : V \to W$ be our linear transformation. Let $v \in V$ then $T(v)$ is said to be the **image** of v under T. Furthermore, if X is any subset of V, then

$$T[X] = \{T(v) : v \in X\}$$

In particular, $T[V]$ is said to be the **range** of T. It is denoted by $\operatorname{ran}(T)$. Similarly if Y is a subset of W then

$$T^{-1}[Y] = \{x \in V : T(x) \in Y\}$$

is the **inverse image of Y under T.**

Remark 2.4.1. Note that if Y is a subspace of W then $T^{-1}[Y]$ may never be the empty set, since Y contains 0_W. Also, $\operatorname{ran}(T) = T[V] \subseteq W$.

We make some important observations:

Theorem 2.4.1. *Let $T : V \to W$ be a linear transformation from V to W. The following holds:*

1. *For all $x, y \in V$, $T(x - y) = T(x) - T(y)$.*
2. *T maps the additive identity $\underline{0}_V$ of V to the additive identity $\underline{0}_W$ of W.*
3. *If $T' : W \to Z$ is also a linear transformation, their composition $T' \circ T$ is also a linear transformation from V to Z where $T' \circ T(v) = T'(T(v))$*
4. *If V_1 is a subspace of V and W_1 is a subspace of W, then $T[V_1]$ is a subspace of W and $T^{-1}[W_1]$ is a subspace of V.*

Exercise 2.4.1. Prove **Theorem 2.4.1**

Remark 2.4.2. The kernel, $\ker(T)$, of a linear transformation $T : V \to W$, which is defined to be

$$\ker(T) = \{v \in V : T(v) = \underline{0}_W\}$$

is a subspace of V

Kernels of linear transformation allows us to determine whether or not a linear map is injective. It also allows us to form a certain **direct sum** of a vector space as we will see soon.

Theorem 2.4.2. *$T : V \to W$ is* ***injective*** $\iff$ $\ker T = \{\underline{0}_V\}$

Exercise 2.4.2. Prove **Theorem 2.4.2**

Remark 2.4.3. If $T : V \to W$ is a linear transformation and $|V| > |W|$ then T can never be injective.

Remark 2.4.4. If $T : V \to W$ is a linear transformation and $|V| < |W|$ then T can never be surjective. That is $T[V] \subset W$

Example 2.4.1. Let $T : \mathbb{P}_3[x] \to \mathbb{P}_3[x]$ be defined by $T(p(x)) = p'(x)$. Clearly differentiation is a linear map and $\ker T = \mathrm{sp}\{I(x) = 1\}$. So, it is the subspace of **constant** functions. Note that $\ker T \subset \ker T^2$ where $T^2 = T \circ T$. T^2 just represents the second derivative. Hence, $\ker T^2 = \mathrm{sp}\{1, x\}$. Since the constant function 1 and the linear function x in $\mathbb{P}_3[x]$ are linearly independent, $\mathcal{B} = \{1, x\}$ is a basis for $\ker T^2$. However, note that $\mathrm{ran}(T^2) \subset \mathrm{ran}(T) = \mathrm{sp}\{1, x, x^2\}$. This example can illustrate another important fact that if v_1 and v_2 are two elements in a vector space that do not lie in a given subspace, their **vector addition** may lie in that particular subspace. Take $v_1 = \sqrt{2}x^3 - 6$ and $v_2 = \sqrt{2}x^3 - 3$. Clearly neither v_1 nor v_2 is in $\ker T$ but $v_1 - v_2 = -3 \in \ker T$

Now let us see a different type of vector spaces. Given two vector spaces V and W over the same field $\mathbb{F}$, let $\mathcal{L}(V, W) = \{T |\ T : V \to W \text{ where } T \text{ is a linear transformation}\}$

Theorem 2.4.3. $(\mathcal{L}(V, W), \boxplus, \odot, (\mathbb{F}, +, \cdot))$ *forms a vector space over* $\mathbb{F}$*, where* $\boxplus$ *and* $\odot$ *are the pointwise addition of linear maps and scalar multiplication with the linear map in* W *respectively.*
That is given $T, S \in \mathcal{L}(V, W)$ *and* $\alpha \in \mathbb{F}$

$$T \boxplus S := T \underbrace{+}_{W} S$$

$$\alpha \odot T := \alpha \underbrace{\cdot}_{W} T$$

So, given any $v \in V$ *the image of* v *under the linear maps* $T \boxplus S$ *and* $\alpha \odot T$ *would be* $T(v) +_W S(v)$ *and* $\alpha \cdot_W T(v)$ *respectively.*

Exercise 2.4.3. Prove **Theorem 2.4.3**. You need to basically prove that $\mathcal{L}(V, W)$ is closed under both $\boxplus$ and $\odot$. That is to show that $T \boxplus S$ and $\alpha \odot T$ are indeed linear transformations from V to W and all the axioms of a vector space are satisfied.

Now there are many vector spaces that possess the same algebraic structures. We call such spaces **isomorphic vector spaces**.

Definition 2.4.4. (Isomorphism) $T \in \mathcal{L}(V, W)$ is called an **isomorphism** if T is injective and surjective.

Remark 2.4.5. Two vector spaces are isomorphic if they have the same algebraic properties. Hence, if there is an isomorphism between two vector spaces, then any algebraic property of one vector in one of the vector spaces that can be obtained from the axioms corresponds to the identical algebraic property of the vector in the other vector space.

Example 2.4.2. Let V and W be two vector spaces over $\mathbb{F}$ with $\dim(V) = n$ and $\dim(W) = m$. Then $\dim(\mathcal{L}(V, W)) = mn$

Definition 2.4.5 (Invertible Linear Transformation)**.** $T \in \mathcal{L}(V, W)$ is an **invertible** linear map if there exists $S \in \mathcal{L}(W, V)$ such that $S \circ T$ is the identity linear map on V and $T \circ S$ is the identity linear map on W.

Exercise 2.4.4. Prove that a linear map $T : V \to W$ is invertible $\iff$ T is injective and surjective.

Exercise 2.4.5. Let V and W be two vector spaces over a field $\mathbb{F}$ with dimensions n and $m > 3$ respectively. Construct three different linearly independent linear maps from V to W

2.5 Coordinatization and Change of Basis

Before we begin note that given a field we have already deduced that $\mathbb{F}$ can be viewed as a vector space over itself, and as a matter of fact we can even increase the dimension of our vector space. Consider $\mathbb{F}^n = \underbrace{\mathbb{F} \times \cdots \times \mathbb{F}}_{n \text{ times}}$ which is the set of all n-tuple of elements of $\mathbb{F}$. This also can be viewed as a vector space over $\mathbb{F}$, but this time would be of dimension n as opposed to dimension 1. Vector addition and scalar multiplication are the usual componentwise scalar addition and scalar multiplication distributed to all the components.

Now the notion of coordinatization enables us to show that any finite dimensional vector space V over a field $\mathbb{F}$, say of dimension n, will be isomorphic to $\mathbb{F}^n$. Coordinatization hence, will allow us to determine algebraic conditions of vectors of from their "corresponding vectors", **coordinate vectors**, which always lie in $\mathbb{F}^n$. For instance, if we are given an arbitrary set of vectors and we are to determine whether the set is independent; the process of setting up equations, and solving for the scalars and trying to deduce that the scalars are just the zero element of the field $\mathbb{F}$ might be cumbersome and frustrating. Hence, what we do is we look at the coordinate vectors of those particular vectors. Now if they are linearly independent or possess other algebraic properties, the same properties are also satisfied by the vectors these coordinate vectors identity.

Now everything can be summarized by the following statements. It will help us identify the matrix representation associated to a linear transformation relative to some ordered bases. Recall from $A22$ that any matrix $\hat{A}_{m\times n} \in \mathcal{M}_{m\times n}(\mathbb{F})$ induces a transformation, T_A, from an n-dimensional vector space V to an m-dimensional vector space W both over $\mathbb{F}$, specifically through $T_A(v) = A[v]_{\mathcal{B}}$ where $\mathcal{B}$ is an ordered basis of V. Right now a basis for W does not matter.

Definition 2.5.1 (Coordinate Vector)**.** Let V be a vector space of dimension n over a field $\mathbb{F}$ and

$$\mathcal{B} = \langle b_1, b_2, \ldots, b_n\rangle$$

be an ordered basis. Given a vector v, there exists a unique set of scalars $\alpha_1, \ldots, \alpha_n \in \mathbb{F}$ such that

$$v = \alpha_1 b_1 + \alpha_2 b_2 + \cdots + \alpha_n b_n$$

Then $[v]_{\mathcal{B}}$ is called the **coordinate vector** of v relative to the ordered basis $\mathcal{B}$, and

$$[v]_{\mathcal{B}} = (\alpha_1, \ldots, \alpha_n)$$

Remark 2.5.1. If $\mathcal{B}' = \langle b'_1, \ldots, b'_n\rangle$ is another ordered basis to V then there exists unique scalars $\gamma_i \in \mathbb{F}$ such that

$$v = \gamma_1 b'_1 + \gamma_2 b'_2 + \cdots + \gamma_n b'_n$$

So we get that

$$[v]_{\mathcal{B}'} = (\gamma_1, \gamma_2, \ldots, \gamma_n) \neq (\alpha_1, \alpha_2, \ldots, \alpha_n)$$

and as a matter of fact every scalar γ_i may be completely different from scalars α_i, however, they define the same vector v using basis $\mathcal{B}$ and $\mathcal{B}'$.

Example 2.5.1. Let $V = \mathbb{P}_3[x]$ over $\mathbb{R}$. Consider the ordered basis $\mathcal{B} = \langle 2, x^2, x, 2x^3 \rangle$ and $\mathcal{B}' = \langle x+1, x^3, 2x^2, \sqrt{2} \rangle$, and consider the vector $v = 5x^3 - 4x^2 + x + 3$. Then

$$[v]_{\mathcal{B}} = (\frac{3}{2}, -4, 1, \frac{5}{2})$$

Since

$$v = (\frac{3}{2})2 + (-4)x^2 + (1)x + (2.5)2x^3$$

and

$$[v]_{\mathcal{B}'} = (1, 5, -2, \sqrt{2})$$

Since

$$5x^3 - 4x^2 + x + 3 = (1)(x+1) + (5)x^3 + (-2)(2x^2) + (\sqrt{2})(\sqrt{2})$$

Now let us revisit example 2.5.1 and solve it more systematically. First let us consider the basis $\mathcal{S} = \langle x^3, x^2, x, 1 \rangle = \langle s_1, s_2, s_3, s_4 \rangle$ as the standard basis. That means that since $\mathbb{P}_3[x] \cong \mathbb{R}^4$, x^3 gets mapped to the vector $(1,0,0,0)$, $x^2 \longmapsto (0,1,0,0)$, $x \longmapsto (0,0,1,0)$ and lastly the function 1 gets mapped to $(0,0,0,1)$.

Now we have two other ordered basis for $\mathbb{P}_3[x]$ in example 2.5.1, namely $\mathcal{B} = \langle 2, x^2, x, 2x^3 \rangle$ and $\mathcal{B}' = \langle x+1, x^3, 2x^2, \sqrt{2} \rangle$, and we have our vector $v = 5x^3 - 4x^2 + x + 3$. Clearly $[v]_{\mathcal{S}} = (5, -4, 1, 3)$.

If I want to compute $[v]_{\mathcal{B}}$ or $[v]_{\mathcal{B}'}$, I need to write each element of my standard basis in terms of elements of $\mathcal{B}$ and $\mathcal{B}'$ respectively.

In doing so, we should obtain

$$\begin{aligned}
x^3 &= 0 \cdot 2 + 0 \cdot x^2 + 0 \cdot x + \frac{1}{2}(2x^3) \\
x^2 &= 0 \cdot 2 + 1 \cdot x^2 + 0 \cdot x + 0 \cdot (2x^3) \\
x &= 0 \cdot 2 + 0 \cdot x^2 + 1 \cdot x + 0 \cdot (2x^3) \\
1 &= \frac{1}{2} \cdot 2 + 0 \cdot x^2 + 0 \cdot x + 0 \cdot (2x^3)
\end{aligned}$$

Hence $M_{\mathcal{SB}} = \begin{bmatrix} | & | & | & | \\ [s_1]_{\mathcal{B}} & [s_2]_{\mathcal{B}} & [s_3]_{\mathcal{B}} & [s_4]_{\mathcal{B}} \\ | & | & | & | \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & \frac{1}{2} \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \frac{1}{2} & 0 & 0 & 0 \end{bmatrix}$.

Now note $M_{\mathcal{SB}}[v]_{\mathcal{S}} = \begin{bmatrix} 0 & 0 & 0 & \frac{1}{2} \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \frac{1}{2} & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 5 \\ -4 \\ 1 \\ 3 \end{bmatrix} = \begin{bmatrix} \frac{3}{2} \\ -4 \\ 1 \\ \frac{5}{2} \end{bmatrix} = [v]_{\mathcal{B}}$.

Similarly,

$$\begin{aligned}
x^3 &= 0 \cdot (x+1) + 1 \cdot (x^3) + 0 \cdot (2x^2) + 0 \cdot \sqrt{2} \\
x^2 &= 0 \cdot (x+1) + 0 \cdot (x^3) + \frac{1}{2} \cdot (2x^2) + 0 \cdot (\sqrt{2}) \\
x &= 1 \cdot (x+1) + 0 \cdot (x^3) + 0 \cdot (2x^2) + (-\frac{1}{\sqrt{2}}) \cdot (\sqrt{2}) \\
1 &= 0 \cdot (x+1) + 0 \cdot (x^3) + 0 \cdot (2x^2) + \frac{1}{\sqrt{2}} \cdot (\sqrt{2})
\end{aligned}$$

So $M_{\mathcal{S}\mathcal{B}'} = \begin{bmatrix} | & | & | & | \\ [s_1]_{\mathcal{B}'} & [s_2]_{\mathcal{B}'} & [s_3]_{\mathcal{B}'} & [s_4]_{\mathcal{B}'} \\ | & | & | & | \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$.

Now $M_{\mathcal{S}\mathcal{B}'}[v]_{\mathcal{S}} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 5 \\ -4 \\ 1 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 \\ 5 \\ -2 \\ \sqrt{2} \end{bmatrix} = [v]_{\mathcal{B}'}$.

But why does this make sense? Consider two arbitrary ordered basis of a finite dimensional vector space V, $\mathcal{C} = \langle c_1, \dots c_n \rangle$ and $\mathcal{C}' = \langle c'_1, \dots c'_n \rangle$.

Now consider $\mathcal{C}$ then given a vector we know $\exists\,!$ (there exists unique) scalars $\alpha_i \in \mathbb{F}$ such that $v = \sum_{i=1}^{n} \alpha_i c_i$. Now $\mathcal{C}'$ is a basis for V as well. Hence, for any i where $1 \leq i \leq n$ there exists other unique scalars (at least one is distinct) $\gamma_{ij} \in \mathbb{F}$ such that

$$c_i = \sum_{j=1}^{n} \gamma_{ji} c'_j, \text{ so for instance } c_1 = \sum_{j=1}^{n} \gamma_{j1} c'_j$$

$$\text{So } v = \alpha_1 \sum_{j=1}^{n} \gamma_{j1} c'_j + \alpha_2 \sum_{j=1}^{n} \gamma_{j2} c'_j + \cdots + \alpha_n \sum_{j=1}^{n} \gamma_{jn} c'_j \tag{1}$$

But this is captured with the matrix

$$\begin{bmatrix} \gamma_{11} & \gamma_{12} & \cdots & \gamma_{1n} \\ \gamma_{21} & \gamma_{22} & \cdots & \gamma_{2n} \\ \gamma_{31} & \gamma_{32} & \cdots & \gamma_{3n} \\ \vdots & \vdots & & \vdots \\ \gamma_{n1} & \gamma_{n2} & \cdots & \gamma_{nn} \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \vdots \\ \alpha_n \end{bmatrix} = \begin{bmatrix} | & | & & | \\ [c_1]_{\mathcal{C}'} & [c_2]_{\mathcal{C}'} & \dots & [c_n]_{\mathcal{C}'} \\ | & | & & | \end{bmatrix} [v]_{\mathcal{C}} = [v]_{\mathcal{C}'} = \begin{bmatrix} \gamma_{11}\alpha_1 + \gamma_{12}\alpha_2 + \cdots + \gamma_{1n}\alpha_n \\ \gamma_{21}\alpha_1 + \gamma_{22}\alpha_2 + \cdots + \gamma_{2n}\alpha_n \\ \vdots \\ \gamma_{n1}\alpha_1 + \gamma_{n2}\alpha_2 + \cdots + \gamma_{nn}\alpha_n \end{bmatrix}$$

Note that we can rearrange (1):

$$v = (\alpha_1\gamma_{11} + \alpha_2\gamma_{12} + \cdots + \alpha_n\gamma_{1n})c'_1 + \cdots + (\alpha_1\gamma_{n1} + \alpha_2\gamma_{n2} + \cdots + \alpha_n\gamma_{nn})c'_n$$

Now we return to example 2.5.1.

We could have computed $M_{\mathcal{B}\mathcal{B}'} = \begin{bmatrix} | & & | \\ [b_1]_{\mathcal{B}'} & \dots & [b_n]_{\mathcal{B}'} \\ | & & | \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 2 \\ 0 & \frac{1}{2} & 0 & 0 \\ \sqrt{2} & 0 & -\frac{1}{\sqrt{2}} & 0 \end{bmatrix}$,

or $M_{\mathcal{B}'\mathcal{B}} = \begin{bmatrix} | & & | \\ [b'_1]_{\mathcal{B}} & \dots & [b'_n]_{\mathcal{B}} \\ | & & | \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & 0 & 0 & \frac{1}{\sqrt{2}} \\ 0 & 0 & 2 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 \end{bmatrix}$.

Now $M_{\mathcal{B}\mathcal{B}'}[v]_{\mathcal{B}} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 2 \\ 0 & \frac{1}{2} & 0 & 0 \\ \sqrt{2} & 0 & -\frac{1}{\sqrt{2}} & 0 \end{bmatrix} \begin{bmatrix} \frac{3}{2} \\ -4 \\ 1 \\ \frac{5}{2} \end{bmatrix} = \begin{bmatrix} 1 \\ 5 \\ -2 \\ \sqrt{2} \end{bmatrix} = [v]_{\mathcal{B}'}$.

Now note $M_{\mathcal{B}\mathcal{S}}[v]_{\mathcal{B}} = M_{\mathcal{B}'\mathcal{S}}[v]_{\mathcal{B}'}$. Both of these matrices are invertible, since their columns are just the coordinate vectors relative to some **basis** and hence the columns of the matrix would form linearly independent set since any isomorphism from $\mathbb{P}_3[x]$ to $\mathbb{R}^4$ preserves the algebraic properties of vectors.

So $(M_{\mathcal{B}'\mathcal{S}})^{-1} M_{\mathcal{B}\mathcal{S}}[v]_{\mathcal{B}} = [v]_{\mathcal{B}'}$, but note that $(M_{\mathcal{B}'\mathcal{S}})^{-1} = M_{\mathcal{S}\mathcal{B}'}$, so $M_{\mathcal{S}\mathcal{B}'}M_{\mathcal{B}\mathcal{S}}[v]_{\mathcal{B}} = [v]_{\mathcal{B}'}$.

Hence, we can conclude that $M_{\mathcal{B}\mathcal{B}'} = M_{\mathcal{S}\mathcal{B}'}M_{\mathcal{B}\mathcal{S}}$, which illustrates another ways of finding the change of the coordinate matrix.

Definition 2.5.2 (Change-of-Coordinates-Matrix)**.** The matrix $M_{\mathcal{B}\mathcal{B}'}$ is called the **change of coordinate matrix** from the ordered basis $\mathcal{B}$ to the ordered basis $\mathcal{B}'$ of a vector space V over a field $\mathbb{F}$.

Remark 2.5.2. Note we have not said anything about a linear transformation above or mapping vectors to a different vector space. We were just discussing how to represent vectors in a **given** vector space in different coordinate vectors.

Now the following theorem allows us to elaborate on the comment we made at the beginning of this subsection, specifically how can we establish connections between the vector space $\mathbb{F}^n$ and an n-dimensional vector space and from there we establish matrix representation of a linear transformation from a vector space V to a vector space W. Note W could be V itself!

Theorem 2.5.1. *Any n-dimensional vector space V over a field $\mathbb{F}$ is isomorphic to $\mathbb{F}^n$.*

Proof. Let $v \in V$ where V is an n-dimensional vector space with an ordered basis $\mathcal{B} = \langle b_1, \dots, b_n \rangle$. Define the map $\phi : V \to \mathbb{F}^n$ by $v \longmapsto [v]_{\mathcal{B}}$.
We claim that ϕ is an **isomorphism!** ϕ is called the coordinate isomorphism

First note that since $\mathcal{B}$ is a basis for V, every vector v can be expressed **uniquely** as a linear combination of the basis vectors in $\mathcal{B}$. Hence, the coordinate vectors of two distinct vectors v, w in V are completely different. This establishes that the map is injective.

To see what this map is surjective, consider any n-tuple in $\mathbb{F}^n$, say $(\alpha_1, \dots, \alpha_n)$. Then for any $1 \le i \le n$ we have that $\alpha_i b_i \in V$ since $b_i \in V$ and V is closed under scalar multiplication. But V is also closed under vector addition. Hence, the vector $\alpha_1 b_1 + \cdots + \alpha_n b_n$ is the pre-image of the n-tuple $(\alpha_1, \dots, \alpha_n)$ under ϕ.

We only need to show that the map is a linear transformation. Consider any two arbitrary vectors v, w in V. We proceed to show that $\phi(v + w) = [v + w]_\beta = \phi(v) + \phi(w) = [v]_{\mathcal{B}} + [w]_{\mathcal{B}}$.
Since $\mathcal{B}$ is a basis for V there exists unique scalars $\alpha_1, \dots, \alpha_n, \gamma_1, \dots, \gamma_n \in \mathbb{F}$ such that

$$v = \sum_{i=1}^{n} \alpha_i b_i \quad w = \sum_{i=1}^{n} \gamma_i b_i$$

By rearranging terms since vector addition is commutative and scalar multiplication is distributive we get that $v + w = \sum_{i=1}^{n} (\alpha_i + \gamma_i) b_i$. Hence, $\phi(v + w) = (\alpha_1 + \gamma_1, \dots, \alpha_n + \gamma_n)$. However, $(\alpha_1 + \gamma_1, \dots, \alpha_n + \gamma_n) = (\alpha_1, \dots, \alpha_n) + (\gamma_1, \dots, \gamma_n) = [v]_{\mathcal{B}} + [w]_{\mathcal{B}} = \phi(v) + \phi(w)$ ■

The following example will show how the concept of coordinatization helps us conclude certain algebraic properties of a set of vectors in a finite dimensional vector space.

Example 2.5.2. Consider the following linearly independent set $\{1, \sin x, \sin 2x, \dots, \sin nx\} \subset \mathcal{F}(\mathbb{R})$ over $\mathbb{R}$. Find a basis for W, the subspace of $\mathcal{F}(\mathbb{R})$ spanned by the functions:

$$\begin{aligned} f_1(x) &= 3 - \sin x + 3 \sin 2x - \sin 3x + 5 \sin 4x \\ f_2(x) &= 1 + 2 \sin x + 4 \sin 2x - \sin 4x \\ f_3(x) &= -1 + 5 \sin x + 5 \sin 2x + \sin 3x - 7 \sin 4x \\ f_4(x) &= 3 \sin 2x - \sin 4x \end{aligned}$$

Answer: By remark 2.3.3, the size of a spanning set is always bigger than or equal to the size of a linearly independent set. Since the subspace is said to be spanned by $f_1(x), f_2(x), f_3(x)$ and $f_4(x)$, we need to determine whether they are linearly independent. Now it is hard to tackle this question with the usual definition of linear independence. Hence, we use the change of coordination and the fact that $\mathrm{sp}(1, \sin x, \sin 2x, \sin 3x, \sin 4x)$ is a subspace that will contain the subspace spanned by $f_1, \dots, f_4$.

Consider the ordered basis $\mathcal{B} = \langle 1, \sin x, \sin 2x, \sin 3x, \sin 4x \rangle$, Then

$$[f_1(x)]_{\mathcal{B}} = [3, -1, 3, -1, 5]$$
$$[f_2(x)]_{\mathcal{B}} = [1, 2, 4, 0, -1]$$
$$[f_3(x)]_{\mathcal{B}} = [-1, 5, 5, 1, -7]$$
$$[f_4(x)]_{\mathcal{B}} = [0, 0, 3, 0, -1]$$

Reducing the matrix $\begin{bmatrix} 3 & 1 & -1 & 0 \\ -1 & 2 & 5 & 0 \\ 3 & 4 & 5 & 3 \\ -1 & 0 & 1 & 0 \\ 5 & -1 & -7 & -1 \end{bmatrix}$ to the reduced row echelon form we obtain $\begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$.

Since the 1^{st}, 2^{nd}, and 4^{th} columns have pivots in them, they are linearly independent, which is equivalent to saying $f_1(x)$, $f_2(x)$, and $f_4(x)$ are linearly independent.

$\therefore \mathcal{B}_W = \langle f_1(x), f_2(x), f_4(x) \rangle$ is a basis for W.

2.6 Matrix Representation and Change of Basis

Theorem 2.6.1. *[Matrix Representation of a Linear Transformation] Let V and W be finite dimensional vector spaces over a field $\mathbb{F}$ take it to be $\mathbb{R}$ for simplicity, of dimension n and m respectively. Let $\mathcal{C} = \langle c_1, \dots, c_n \rangle$ and $\mathcal{C} = \langle c'_1, \dots, c'_m \rangle$ be the corresponding ordered basis. Let $T \in \mathcal{L}(V, W)$ and $\overline{T} \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ such that*

$$[T(v)]_{\mathcal{C}'} = \overline{T}([v]_{\mathcal{C}})$$

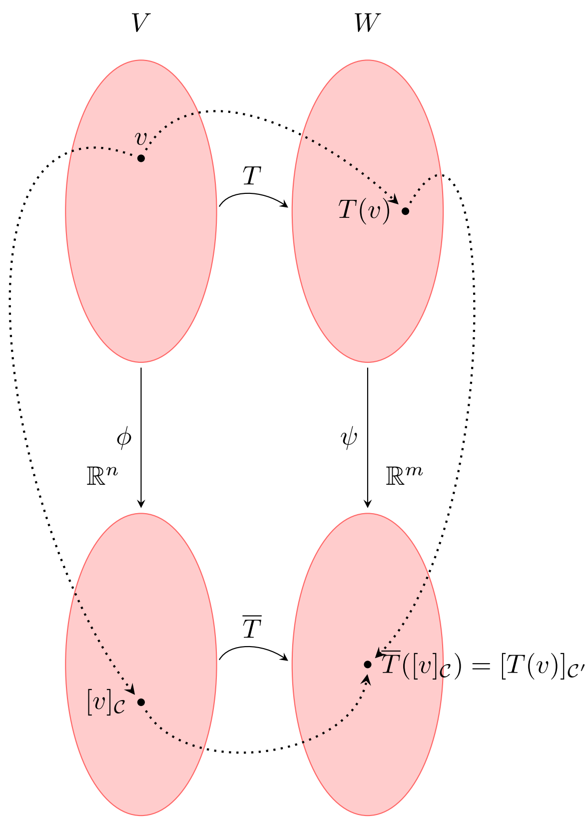
*for all $v \in V$. Then the **matrix representation** of T relative to $\mathcal{C}, \mathcal{C}'$ is the matrix $R_{\mathcal{C}\mathcal{C}'}$ whose j^{th} column is the column vector $[T(c_j)]_{\mathcal{C}'}$ and satisfies:*

$$[T(v)]_{\mathcal{C}'} = R_{\mathcal{C}\mathcal{C}'}[v]_{\mathcal{C}} \qquad (1)$$

for all $v \in V$.

Note that (1) gives us the coordinate vector of the image of an arbitrary vector v under T relative to the ordered basis $\mathcal{C}'$, not the actual image of v under T. However, the image of v can easily by computed by the pull-back map, ψ^{-1}, as will be explained on the next page.

$$R_{\mathcal{C}\mathcal{C}'} = \begin{bmatrix} | & & | \\ [T(c_1)]_{\mathcal{C}'} & \dots & [T(c_n)]_{\mathcal{C}'} \\ | & & | \end{bmatrix}_{m \times n}$$



Proof. Let us first establish the existence of such $\overline{T}$, which allows for the above diagram to commute. Let us explicitly define it. Given any $\mathbf{x} \in \mathbb{R}^{\mathbf{n}}$ we define $\overline{T}$ as follows: First note that since ϕ is an isomorphism then its inverse ϕ^{-1} is also an isomorphism! Let $\mathbf{x} = (\alpha_{\mathbf{1}}, \ldots, \alpha_{\mathbf{n}})$. Then $\phi^{-1}(\mathbf{x}) = \alpha_1 c_1 + \cdots + \alpha_n c_n \in V$. Call $\phi^{-1}(\mathbf{x}) = z$. Apply the coordinate isomorphism ψ on $T(z)$ and we have $\psi(T(z)) = \left[T(z)\right]_{\mathcal{C}'}$.
So define $\overline{T}(\mathbf{x}) := \psi \circ T \circ \phi^{-1}(\mathbf{x}) = \left[T(z)\right]_{\mathcal{C}'}$.
Now let us show why $R_{\mathcal{C}\mathcal{C}'}$ has the form it does and why for any $v \in V$ we have that $T[(v)]_{\mathcal{C}'} = R_{\mathcal{C}\mathcal{C}'}[v]_{\mathcal{C}}$ even though it should be self explanatory from the definition above.

Let $\mathcal{C} = \{c_1, \ldots, c_n\}$ and $\mathcal{C}' = \{c'_1, \ldots, c'_m\}$ be two ordered bases for V and W respectively. Let $v \in V$, then $v = \displaystyle\sum_{i=1}^{n} \alpha_i c_i$ for some unique scalars $\alpha_i \in \mathbb{F}$.

Then by linearity of T, $T(v) = T\left(\displaystyle\sum_{i=1}^{n} \alpha_i c_i\right) = \displaystyle\sum_{i=1}^{n} \alpha_i T(c_i)$.

Now for every $1 \leq i \leq n$, $T(c_i) \in W$, there exist unique scalars γ_{ji}, $1 \leq j \leq m$ such that $T(c_i) = \displaystyle\sum_{j=1}^{m} \gamma_{ji} c'_j$. Note that $T[c_i]_{\mathcal{C}'} = (\gamma_{1i}, \gamma_{2i}, \dots, \gamma_{mi})$

$$\begin{aligned}\text{Now } T(v) &= \sum_{i=1}^{n} \alpha_i \left(\sum_{j=1}^{m} \gamma_{ji} c'_j \right) \\ &= \alpha_1 \left(\gamma_{11} c'_1 + \gamma_{21} c'_2 + \cdots + \gamma_{m1} c'_m \right) + \cdots + \alpha_n \left(\gamma_{1n} c'_1 + \gamma_{2n} c'_2 + \cdots + \gamma_{mn} c'_m \right) \\ &= \left(\alpha_1 \gamma_{11} + \alpha_2 \gamma_{12} + \cdots + \alpha_n \gamma_{1n} \right) c'_1 + \cdots + \left(\alpha_1 \gamma_{m1} + \alpha_2 \gamma_{m2} + \cdots + \alpha_n \gamma_{mn} \right) c'_m\end{aligned}$$

$$\text{So } [T(v)]_{\mathcal{C}'} = \begin{bmatrix} \alpha_1 \gamma_{11} + \alpha_2 \gamma_{12} + \cdots + \alpha_n \gamma_{1n} \\ \alpha_1 \gamma_{21} + \alpha_2 \gamma_{22} + \cdots + \alpha_n \gamma_{2n} \\ \vdots \\ \alpha_1 \gamma_{m1} + \alpha_2 \gamma_{m2} + \cdots + \alpha_n \gamma_{mn} \end{bmatrix}.$$

Now consider the following matrix:

$$[T(v)]_{\mathcal{C}'} = \begin{bmatrix} \gamma_{11} & \gamma_{12} & \cdots & \gamma_{1n} \\ \gamma_{21} & \gamma_{22} & \cdots & \gamma_{2n} \\ \gamma_{31} & \gamma_{32} & \cdots & \gamma_{3n} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{m1} & \gamma_{m2} & \cdots & \gamma_{mn} \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \vdots \\ \alpha_n \end{bmatrix} = \begin{bmatrix} | & & | \\ [T(c_1)]_{\mathcal{C}'} & \cdots & [T(c_n)]_{\mathcal{C}'} \\ | & & | \end{bmatrix} [v]_{\mathcal{C}} = \mathcal{R}_{\mathcal{C}\mathcal{C}'} [v]_{\mathcal{C}}$$

■

Example 2.6.1. Consider the following linear transformation:

$$T \in \mathcal{L}(\mathbb{P}_2[x], \mathbb{P}_3[x]) \tag{3}$$
$$T(p(x)) = (x+1)p(x-2) \tag{4}$$

Find the matrix representation $R_{\mathcal{B}\mathcal{B}'}$ of T relative to the ordered basis

$$\mathcal{B} = \langle x^2, x, 1\rangle \quad \mathcal{B}' = \langle x^3, x^2, x, 1\rangle$$

Also find the image of $p(x) = 5x^2 - 7x + 18$
SOL: We compute the images of the elements of $\mathcal{B}$ and use their coordinate vectors with respect to $\mathcal{B}'$ to set up our matrix representation:

$$\begin{aligned} T(x^2) &= x^3 - 3x^2 + 4 & & & [T(x^2)]_{\mathcal{B}'} &= (1, -3, 0, 4) \\ T(x) &= (x+1)(x-2) = x^2 - x - 2 & &\Longrightarrow & [T(x)]_{\mathcal{B}'} &= (0, 1, -1, -2) \\ T(1) &= x + 1 & & & [T(1)]_{\mathcal{B}'} &= (0, 0, 1, 1) \end{aligned}$$

So we get

$$R_{\mathcal{B}\mathcal{B}'} = \begin{bmatrix} 1 & 0 & 0 \\ -3 & 1 & 0 \\ 0 & -1 & 1 \\ 4 & -2 & 1 \end{bmatrix}$$

Using this, we can compute the image of $p(x) = 5x^2 - 7x + 18$ as follows:

$$R_{\mathcal{B}\mathcal{B}'}[p(x)]_{\mathcal{B}} = \begin{bmatrix} 1 & 0 & 0 \\ -3 & 1 & 0 \\ 0 & -1 & 1 \\ 4 & -2 & 1 \end{bmatrix} \begin{bmatrix} 5 \\ -7 \\ 18 \end{bmatrix} = \begin{bmatrix} 5 \\ -22 \\ 25 \\ 52 \end{bmatrix}$$

So we get that $T(p(x)) = 5x^3 - 22x^2 + 25x + 52$.

Example 2.6.2. Let V be the subspace $\text{sp}\big\{\sin(x)\cos(x), \sin^2(x), \cos^2(x)\big\}$ of the vector space of all differentiable functions, $\mathbf{D}(\mathbb{R})$. Consider the following linear transformation:

$$\begin{aligned} T : V &\to V \\ f(x) &\longmapsto f'(x) \end{aligned}$$

Find the matrix representation of T relative to $\mathcal{B}$ where

$$\mathcal{B} = \big\langle \sin(x)\cos(x), \sin^2(x), \cos^2(x) \big\rangle$$

Compute the derivative of $f(x) = 3\sin(x)\cos(x) - 5\sin^2(x) + 7\cos^2(x)$.
SOL: We set up the matrix representation of T by first computing the following:

$$\begin{aligned} T(\sin(x)\cos(x)) &= -\sin^2(x) + \cos^2(x) & & & [T(\sin(x)\cos(x))]_{\mathcal{B}} &= (0, -1, 1) \\ T(\sin^2(x)) &= 2\sin(x)\cos(x) & &\Longrightarrow & [T(\sin^2(x))]_{\mathcal{B}} &= (2, 0, 0) \\ T(\cos^2(x)) &= -2\cos(x)\sin(x) & & & [T(\cos^2(x))]_{\mathcal{B}} &= (-2, 0, 0) \end{aligned}$$

So we get the matrix representation of T as the following:

$$R_{\mathcal{B}} = \begin{bmatrix} 0 & 2 & -2 \\ -1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

Then to find $T(3\sin(x)\cos(x) - 5\sin^2(x) + 7\cos^2(x))$, we note by theorem 2.6.1 that

$$\begin{aligned} [T(3\sin(x)\cos(x) - 5\sin^2(x) + 7\cos^2(x))]_{\mathcal{B}} &= R_{\mathcal{B}}[3\sin(x)\cos(x) - 5\sin^2(x) + 7\cos^2(x)]_{\mathcal{B}} \\ &= \begin{bmatrix} 0 & 2 & -2 \\ -1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 3 \\ -5 \\ 7 \end{bmatrix} \\ &= \begin{bmatrix} -24 \\ -3 \\ 3 \end{bmatrix} \qquad (*) \end{aligned}$$

So $(*)$ is the coordinate vector of the image of the vector

$$v = 3\sin(x)\cos(x) - 5\sin^2(x) + 7\cos^2(x)$$

under differentiation relative to the ordered basis $\mathcal{B}$, hence the actual image of v is

$$T(v) = -24\sin(x)\cos(x) - 3\sin^2(x) + 3\cos^2(x)$$

Example 2.6.3. Let $T \in \mathcal{L}(\mathbb{R}^3, \mathbb{R}^3)$ and $\mathcal{S} = \langle (1,0,0), (0,1,0), (0,0,1) \rangle$ and $\mathcal{B} = \langle (1,1,0), (1,0,1), (0,1,1) \rangle$ be two ordered bases for $\mathbb{R}^3$. Let

$$\mathcal{R}_{\mathcal{S}} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Find $R_{\mathcal{B}}$. Solve it in two different ways.

Solution I:
We can compute $R_{\mathcal{B}} = M_{\mathcal{SB}} R_{\mathcal{S}} M_{\mathcal{BS}}$.

Since we are given $R_{\mathcal{S}}$ and we can easily compute the change of coordinate matrices $M_{\mathcal{BS}}$ and $M_{\mathcal{SB}}$.

$$M_{\mathcal{BS}} = \begin{bmatrix} | & | & | \\ [b_1]_{\mathcal{S}} & [b_2]_{\mathcal{S}} & [b_3]_{\mathcal{S}} \\ | & | & | \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$$

$$M_{\mathcal{SB}} = \begin{bmatrix} | & | & | \\ [s_1]_{\mathcal{B}} & [s_2]_{\mathcal{B}} & [s_3]_{\mathcal{B}} \\ | & | & | \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

$$R_{\mathcal{B}} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

Solution II:

$$R_{\mathcal{B}} = \begin{bmatrix} | & | & | \\ [T(b_1)]_{\mathcal{B}} & [T(b_2)]_{\mathcal{B}} & [T(b_3)]_{\mathcal{B}} \\ | & | & | \end{bmatrix}$$

But T is not given to us. However, using $R_{\mathcal{S}}$, we can find a formula for T.

Pick $x, y, z \in \mathbb{R}^3$.

Note that

$$\mathcal{R}_{\mathcal{S}} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

so

$$[T(1,0,0)]_{\mathcal{S}} = (1,1,0)$$

$$[T(0,1,0)]_{\mathcal{S}} = (1,1,0)$$

$$[T(0,0,1)]_{\mathcal{S}} = (1,0,1)$$

Thus by linearity,

$$\begin{aligned} T(x,y,z) &= T(x(1,0,0) + y(0,1,0) + z(0,0,1)) \\ &= xT(1,0,0) + yT(0,1,0) + zT(0,0,1) \\ &= x(1,1,0) + y(1,1,0) + z(1,0,1) \\ &= (x+y+z, x+y, z) \end{aligned}$$

Hence, we have found the formula for T.

Now we can compute

$$\begin{aligned} [T(b_1)]_{\mathcal{B}} &= [T(1,1,0)]_{\mathcal{B}} = [(2,2,0)]_{\mathcal{B}} = (2,0,0) \\ [T(b_2)]_{\mathcal{B}} &= [T(1,0,1)]_{\mathcal{B}} = [(2,1,1)]_{\mathcal{B}} = (1,1,0) \\ [T(b_3)]_{\mathcal{B}} &= [T(0,1,1)]_{\mathcal{B}} = [(2,1,1)]_{\mathcal{B}} = (1,1,0) \end{aligned}$$

Therefore,

$$R_{\mathcal{B}} = \begin{bmatrix} | & | & | \\ [T(b_1)]_{\mathcal{B}} & [T(b_2)]_{\mathcal{B}} & [T(b_3)]_{\mathcal{B}} \\ | & | & | \end{bmatrix} = \begin{bmatrix} 2 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

Example 2.6.4. Consider the vector space $\mathbb{P}_2[x]$ over $\mathbb{R}$. Define $T : \mathbb{P}_2[x] \to \mathbb{P}_2[x]$ by $p(x) \mapsto p(x-1)$. Consider the ordered bases $\mathcal{S} = \langle x^2, x, 1\rangle$ and $\mathcal{B} = \langle x, x+1, x^2-1\rangle$. Find the matrix representations $R_{\mathcal{B}}$ of T.

Answer: First of all, $T(x^2) = (x-1)^2 = x^2 - 2x + 1, T(x) = x - 1, T(1) = 1$.

Then

$$R_{\mathcal{S}} = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 1 & -1 & 1 \end{bmatrix}$$

Now we compute the change of coordinate matrices $M_{\mathcal{BS}}$ and $M_{\mathcal{SB}}$.

$$M_{\mathcal{BS}} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & -1 \end{bmatrix} \quad \text{and} \quad M_{\mathcal{SB}} = \begin{bmatrix} -1 & 1 & -1 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$$

So,

$$R_{\mathcal{B}} = M_{\mathcal{SB}} R_{\mathcal{S}} M_{\mathcal{BS}} = \begin{bmatrix} -1 & 1 & -1 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 1 & -1 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & -1 \end{bmatrix} = \begin{bmatrix} 2 & 1 & -3 \\ -1 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

Alternatively, we could have computed

$$R_{\mathcal{B}} = \begin{bmatrix} | & | & | \\ [T(x)]_{\mathcal{B}} & [T(x+1)]_{\mathcal{B}} & [T(x^2-1)]_{\mathcal{B}} \\ | & | & | \end{bmatrix} = \begin{bmatrix} | & | & | \\ [x-1]_{\mathcal{B}} & [x]_{\mathcal{B}} & [x^2-2x]_{\mathcal{B}} \\ | & | & | \end{bmatrix} = \begin{bmatrix} 2 & 1 & -3 \\ -1 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

Exercise 2.6.1. Prove Example 2.4.2

Remark 2.6.1. All these examples should demonstrate the effect of changing basis for coordinatization has on the matrix representation of a given linear transformation. But let us try to delve a little bit deeper.

Remark 2.6.2. Let us explore things in more detail. We have already concluded any arbitrary linear map T from V to W where $\dim(V) = n$ and $\dim(W) = m$ yields a matrix representation relative to suitable bases. Now how about if we go the other way around and start off with a random matrix. What will happen then? Can find linear transformation from defined on appropriate spaces whose matrix representation will be the matrix we started off with?

Let $\hat{A} = [a_{ij}]_{m\times n} \in \mathcal{M}_{m\times n}(\mathbb{R})$. Then A will induce a linear transformation $\overline{T}$ from $\mathbb{R}^n$ to $\mathbb{R}^m$ in a natural way. Pick an element $\underline{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$. Then $\overline{T}_A(\underline{x}) = A \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = [a_{ij}]_{m\times n} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^m$. So, it should be clear that if we consider the standard bases $\mathcal{S} = \langle \underbrace{(1, 0, \ldots, 0)}_{\text{n-tuple}}, \ldots, (0, \ldots, 1)\rangle$ and $\mathcal{S}' = \langle \underbrace{(1, 0, \ldots, 0)}_{\text{m-tuple}}, \ldots, (0, \ldots, 1)\rangle$ of $\mathbb{R}^n$ and $\mathbb{R}^m$ respectively, then clearly the matrix representation of $\overline{T}_A$ relative to $\mathcal{S}, \mathcal{S}'$, $R_{\mathcal{S}\mathcal{S}'}$, would be A itself. For instance $(a_{11}, a_{21}, \ldots, a_{m1})$ would be simply $[\overline{T}_A(\underbrace{(1, 0, \ldots, 0)}_{\text{n-tuple}})]_{\mathcal{S}'}$.

In addition, now if we have any two vector spaces V and W of dimension n and m respectively over $\mathbb{R}$ (or any other fields to that matter say $\mathbb{C}$ in which case we would replace $\mathbb{R}^n$ by $\mathbb{C}^n$) with ordered bases $\mathcal{B} = \langle b_1, \ldots, b_n\rangle$ and $\mathcal{B}' = \langle b_1', \ldots, b_m'\rangle$, then $\overline{T}_A$ would also in turn induce a linear map T from V to W using a "pull-back map." So consider the inverse image, or pre-image, of the point $\underline{x} = (x_1, \ldots, x_n)$ under the coordinate isomorphism map induced by $\mathcal{B}$. We obtain $v = \sum_{i=1}^{n} x_i b_i \in V$. Now in order to define the image of v under T, $T(v)$, first consider the vector $A \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ in $\mathbb{R}^m$, say it is equal to $\begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}$ Now apply a pull back map, again that is to consider the pre-image of $(y_1, \ldots, y_m)$ under the coordinate isomorphism induced by $\mathbb{B}'$. Thereby we obtain $\sum_{i=1}^{m} y_i b_i'$, and element of W. Now map $v \longmapsto \sum_{i=1}^{m} y_i b_i'$.

The matrix representation of the coordinate isomorphism we can view it simply as the identity map. We are mapping

$$b_1 \longmapsto (1, 0, \ldots, 0), \ldots, b_n \longmapsto (0, \ldots, 1)$$

Remark 2.6.3. So, we can easily compute the $\ker(T)$ or $\mathrm{Im}(T)$ by considering the matrix representation R of T. So finding the **nullspace** of R would be equivalent to finding the $\ker(T)$ and **column space** of R would be related to finding the $\mathrm{Im}(T)$. Hence, the **rank-nullity theorem** $\mathrm{rank}(R) + \mathrm{null}(R) = \#\mathrm{col}(R)$ can be translated to $\dim(\ker(T)) + \dim(\mathrm{Im}(T)) = \dim(V)$ if $T : V \to W$

Now let us recall the following definition:

Definition 2.6.1 (similar Matrices)**.** Let $\hat{A}$ and $\hat{B}$ be two $n \times n$ matrices. $\hat{A}$ and $\hat{B}$ are called **similar** if there exists an invertible matrix $\hat{C}$ such that $A = C^{-1}BC$

Remark 2.6.4. Let $T \in \mathcal{L}(V)$, where V is any n-dimensional vector space over $\mathbb{F}$. Consider any two ordered bases $\mathcal{B}$ and $\mathcal{B}'$ for V. Then $R_{\mathcal{B}\mathcal{B}}$ or simply $R_{\mathcal{B}}$ and $R_{\mathcal{B}'\mathcal{B}'}$ or simply $R_{\mathcal{B}'}$ are similar matrices through the invertible change-of-coordinate matrix $M_{\mathcal{B}\mathcal{B}'}$

Exercise 2.6.2. Suppose A and B are two similar matrices of size n. Prove whether or not they represent, or can be viewed as matrix representations of the same linear transformation relative to suitable ordered bases

3 Inner Products Over $\mathbb{C}$

3.1 Complex Numbers

Definition 3.1.1 (Complex Number)**.** The set of **Complex Numbers**, $\mathbb{C}$, is the set $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$ equipped with two binary operations $+$ and $\cdot$ where

$$\begin{aligned} + : \mathbb{R}^2 \times \mathbb{R}^2 &\to \mathbb{R}^2 \\ \big((x_1, y_1), (x_2, y_2)\big) &\longmapsto (x_1 + x_2, y_1 + y_2) \\ \cdot : \mathbb{R}^2 \times \mathbb{R}^2 &\to \mathbb{R}^2 \\ \big((x_1, y_1), (x_2, y_2)\big) &\longmapsto (x_1 x_2 - y_1 y_2, x_1 y_2 + y_1 x_2) \end{aligned}$$

Now let $i = \sqrt{-1} \iff i^2 = -1$ then we can view $\mathbb{C} = \{a + ib : a, b \in \mathbb{R}\}$ and translate the operations as follows: $z_1 = a + ib$ and $z_2 = c + id$ then $z_1 + z_2 = \underbrace{(a + c)}_{\mathrm{Re}(z_1+z_2)} + i \underbrace{(b + d)}_{\mathrm{Im}(z_1+z_2)}$ and $z_1 \cdot z_2 = ac - bd + i(ad + bc)$

Definition 3.1.2 (Conjugate)**.** Given $z \in \mathbb{C}$, **conjugate** $\overline{z}$ of $z = a + ib$ is defined to be $a - bi$

Theorem 3.1.1. *The triple* $(\mathbb{C}, +, \cdot)$ *as defined above forms a field.*

Proof. Clearly we only need to establish the existence of multiplicative inverse for any non-zero element. Let $0 \neq z = a + ib \in \mathbb{C}$. Then we claim $z^{-1} = \frac{1}{z}$. But what is $\frac{1}{z}$? It does not look like an element of $\mathbb{C}$. But it is! $\frac{1}{z} \cdot \frac{\overline{z}}{\overline{z}} = \frac{a-ib}{a^2+b^2} = \frac{a}{a^2+b^2} - i\frac{b}{a^2+b^2} \in \mathbb{C}$ ∎

Remark 3.1.1. Hence, $\mathbb{C}$ when viewed as a vector space over $\mathbb{R}$ is a 2-dimensional space. A possible basis would be $\{i, a\}$ where $a \in \mathbb{R} \setminus \{0\}$. Therefore $\mathbb{C}$ as a vector space is isomorphic to the vector space $\mathbb{R}^2$ over $\mathbb{R}$ since they both have the same dimension. That is why we can always view points of complex set in $\mathbb{R}^2$. However, if $\mathbb{C}$ is viewed as a vector space over itself then it would be of dimension 1 where any $\alpha \in \mathbb{C} \setminus \{0\}$ can serve as a basis for it.

Proposition 3.1.1. *Properties of complex conjugation that can be easily verified.*

1. $\overline{z_1 + z_2} = \overline{z}_1 + \overline{z}_2$
2. $\overline{z_1 \cdot z_2} = \overline{z}_1 \cdot \overline{z}_2$
3. $\overline{\frac{z_1}{z_2}} = \frac{\overline{z}_1}{\overline{z}_2}$
4. $\overline{\overline{z}} = z$

Remark 3.1.2. Note that any complex number $z = a + ib$ can be expressed in a **polar form**.
So, $z = r(\cos(\theta) + i\sin(\theta))$

Remark 3.1.3. **Euler's Formula:** For any complex number z, $e^{iz} = \cos(z) + i\sin(z)$ which can easily be derived from the Taylor Series expansion of $e^x = \sum_{n=0} \frac{x^n}{n!}$, $\cos(x) = \sum_{n=0} (-1)^n \frac{x^{2n}}{(2n)!}$ and $\sin(x) = \sum_{n=1} (-1)^n \frac{x^{2n-1}}{(2n-1)!}$ and replace x with ix. Hence, $a + ib = z = r(\cos(\theta) + i\sin(\theta)) = re^{i\theta}$, where $a, b \in \mathbb{R}$ and $r = \sqrt{a^2 + b^2}$ and θ is the angle with the x-axis

Remark 3.1.4. We define the length of a complex number $z = a + bi$ to be $|z| = \sqrt{z\overline{z}} = \sqrt{a^2 + b^2}$. So, $|z|^2 = z\overline{z}$. Note that the map that sends $a + ib$ in the complex plane to the point (a, b) in $\mathbb{R}^2$ is **isometric isomorphism**, a notion that will see soon. Basically means not only it is an isomorphism but preserves the length of vectors. Now note that the Euclidean length of the vector (a, b) in $\mathbb{R}^2$ (our usual notion of length, the smallest **distance** to the origin $(0, 0)$) is exactly $\sqrt{a^2 + b^2}$

Remark 3.1.5. You might ponder are there any differences between $\mathbb{R}$ and $\mathbb{C}$? When viewed as vector spaces, the answer is obvious! One is a 1-dimension vector space over $\mathbb{R}$ the other is a 2-dimensional space. Hence, they are not isomorphic vector spaces. Therefore, there must be plenty of differences in their algebraic structures. But some are very vivid. For instance, in the space $\mathbb{C}$ we have no notion of order whereas in $\mathbb{R}$ we do. We know $\sqrt{2} < 3$. Furthermore, in $\mathbb{R}$ we can classify any non-zero element as positive or negative real number. $x \in \mathbb{R}$ is positive if there exists $\alpha \in \mathbb{R}$ such that $\alpha^2 = x$. x is said to be negative if there is NO such α. In $\mathbb{C}$ we cannot classify positive/negative numbers. "Loosely speaking" think of everything as squared, for instance no notion of $-\infty$. Only ∞

3.2 Introduction

The following mathematical notion allows us to describe the concepts of **length** of a vector and **angle** between two vectors in general vector spaces.

Definition 3.2.1 (Hermitian Inner Product Space)**.** A vector space V over $\mathbb{C}$ is called an **inner product space** if it can be equipped with a function

$$\langle\ ,\rangle : V \times V \to \mathbb{C}$$

with the following properties:

1. $\textbf{Positive} - \textbf{Definiteness}$: For all $v \in V$, $\langle v, v\rangle \geq 0$ and $\langle v, v\rangle = 0$ if and only if $v = 0_V$.
2. **Conjugate Symmetry** : For all $v, w \in V$, $\langle v, w\rangle = \overline{\langle w, v\rangle}$
3. **Additivity in the 2$^{\text{nd}}$ slot** : For all $u, v, w \in V$, $\langle u, v + w\rangle = \langle u, v\rangle + \langle u, w\rangle$
4. **Homogeneity in the 2$^{\text{nd}}$ slot** : For all $v, u \in V$ and $\alpha \in \mathbb{C}$, $\langle u, \alpha v\rangle = \alpha \langle u, v\rangle$

Remark 3.2.1. Properties $(3), (4)$ could be combined and simply referred to as **Linearity in the** 2^{nd} **slot**

Remark 3.2.2. We could have defined linearity in the **first** slot as well. As a matter of fact many authors do for simplicity.

Proposition 3.2.1. *Basic properties of an inner product space V over $\mathbb{C}$*

(I) $\forall v \in V$, $\langle v, 0\rangle = \langle 0, v\rangle = 0_{\mathbb{C}}$

(II) $\forall u, v, w \in V$, $\langle v + u, w\rangle = \langle v, w\rangle + \langle u, w\rangle$

(III) $\forall u, v \in V$, $\forall \alpha \in \mathbb{C}$, $\langle \alpha u, v\rangle = \overline{\alpha}\langle u, v\rangle$

(IV) $\forall v, u \in V$, $\langle u + v, u + v\rangle = \langle u, u\rangle + 2\operatorname{Re}(\langle u, v\rangle) + \langle v, v\rangle$

Proof. We prove properties (I), (III), (IV) and leave (II) as an exercise.

(I) Let $v \in V$, then

$$\begin{aligned}
\langle 0, v\rangle &= \langle 0 + 0, v\rangle \\
&= \overline{\langle v, 0 + 0\rangle} && \text{(conjugate symmetry)} \\
&= \overline{\langle v, 0\rangle + \langle v, 0\rangle} && \text{(linearity in 2nd slot)} \\
&= \overline{\langle v, 0\rangle} + \overline{\langle v, 0\rangle} \quad (1) && \text{(property of conjugate)}
\end{aligned}$$

Note $\langle 0, v\rangle = \overline{\langle v, 0\rangle}$, since $\overline{\langle v, 0\rangle} \in \mathbb{C}$, it has an additive inverse, $-\overline{\langle v, 0\rangle}$. Adding $-\overline{\langle v, 0\rangle}$ to both sides of equation (1) we obtain $\overline{\langle v, 0\rangle} = 0_{\mathbb{C}} = \langle 0, v\rangle$

(III) Let $u, v \in V$ and $\alpha \in \mathbb{C}$, then

$$\begin{aligned}
\langle \alpha u, v\rangle &= \overline{\langle v, \alpha u\rangle} && \text{(conjugate symmetry)} \\
&= \overline{\alpha\langle v, u\rangle} && \text{(homogeneity in the 2nd slot)} \\
&= \overline{\alpha}\overline{\langle v, u\rangle} && \text{(property of conjugate)} \\
&= \overline{\alpha}\langle u, v\rangle && \text{(conjugate symmetry)}
\end{aligned}$$

(IV) Let $u, v \in V$.

$$\begin{aligned}
\langle u + v, u + v\rangle &= \langle u + v, u\rangle + \langle u + v, v\rangle && \text{(linearity in the 2nd slot)} \\
&= \overline{\langle u, u + v\rangle} + \overline{\langle v, u + v\rangle} && \text{(conjugate symmetry)} \\
&= \overline{\langle u, u\rangle + \langle u, v\rangle} + \overline{\langle v, u\rangle + \langle v, v\rangle} && \text{(linearity in the 2nd slot)} \\
&= \overline{\langle u, u\rangle} + \overline{\langle u, v\rangle} + \overline{\langle v, u\rangle} + \overline{\langle v, v\rangle} && \text{(property of conjugate)} \\
&= \langle u, u\rangle + \overline{\langle u, v\rangle} + \langle u, v\rangle + \langle v, v\rangle && \text{(conjugate symmetry)} \\
&= \langle u, u\rangle + \overline{\langle u, v\rangle} + \langle u, v\rangle + \langle v, v\rangle && (\overline{\langle u, v\rangle} + \langle u, v\rangle = 2\operatorname{Re}(\langle u, v\rangle))
\end{aligned}$$

Note that for any $\langle u, v\rangle = a + bi \in \mathbb{C}$, $\overline{\langle u, v\rangle} = a - bi$, so $\langle u, v\rangle + \overline{\langle u, v\rangle} = a + bi + a - bi = 2a = 2\operatorname{Re}(\langle u, v\rangle)$

■

Exercise 3.2.1. Prove property (II).

Example 3.2.1. Let V be an inner product space over $\mathbb{C}$. Then there is a linear map $T \in \mathcal{L}(V, \mathbb{C})$.

Answer: Fix $v \in V$. Then for every $u \in V$ map u to $\langle v, u \rangle$.

$$\therefore T : V \longrightarrow \mathbb{C}$$

$$u \longmapsto \langle v, u \rangle$$

But why is it linear? The linearity of T follows from the fact that the inner product is additive and homogeneous in the 2^{nd} slot.

Remark 3.2.3. Example above can be employed to present another proof of property (I) of inner product! How? Note that any linear map sends the additive identity of its domain to the additive identity of its co-domain.

$$0_V \mapsto \langle v, 0 \rangle = 0_{\mathbb{F}}$$

Example 3.2.2. The vector space $(\mathcal{C}([0,1], \mathbb{R}), +, \cdot, (\mathbb{R}, +, \cdot))$ (continuous real-valued functions on $[0,1]$ with the usual pointwise addition and multiplication of functions) can be equipped with an inner product.
Define

$$\langle\ ,\ \rangle : \mathcal{C}([0,1], \mathbb{R}) \times \mathcal{C}([0,1], \mathbb{R}) \longrightarrow \mathbb{R}$$

$$(f(x), g(x)) \longmapsto \langle f(x), g(x) \rangle = \int_0^1 f(x)g(x)\ dx$$

Exercise 3.2.2. Verify the above function is an inner product on $\mathcal{C}([0,1], \mathbb{R})$

Example 3.2.3. Consider the vector space $\mathbb{C}^n$ over $\mathbb{C}$. Define an inner product on it.

Answer: Given $(\underline{\alpha}, \underline{\beta}) \in \mathbb{C}^n \times \mathbb{C}^n$ where $\underline{\alpha} = (\alpha_1, \ldots, \alpha_n)$ and $\underline{\beta} = (\beta_1, \ldots, \beta_n)$ map $(\underline{\alpha}, \underline{\beta})$ to $\langle \underline{\alpha}, \underline{\beta} \rangle = \sum_{i=1}^{n} \overline{\alpha}_i \beta_i$.

Exercise 3.2.3. Verify the above function is an inner product on $\mathbb{C}^n$

Exercise 3.2.4. Could we have defined $\langle \underline{\alpha}, \underline{\beta} \rangle = \sum_{i=1}^{n} \alpha_i \overline{\beta}_i$. Would it be an inner product on $\mathbb{C}^n$?

Remark 3.2.4. Note the map $\langle \underline{\alpha}, \underline{\beta} \rangle = \sum_{i=1}^{n} \alpha_i \beta_i$ would not be an **inner product** on $\mathbb{C}^n$. Why? Positive-definiteness would not be satisfied. For simplicity take $n = 2$ and take $\alpha = (1, i) \in \mathbb{C}^2$. Then $\langle (1, i), (1, i) \rangle = 1.1 + i.i = 1 + i^2 = 0$. However, $(1, i)$ is not the additive zero of the vector space $\mathbb{C}^2$.

The following notion captures the concept of the "length" of a vector in a vector space.

Definition 3.2.2 (Norm)**.** Let V be a vector space. A **norm** on V is a function $\|\cdot\| : V \to [\,0,\infty)$ with the following properties:

1. $\textbf{Positive}-\textbf{Definiteness}$: For all $v \in V$ $\|v\| = 0 \iff v = 0_V$.
2. **Homogeneity** : For all $v \in V$ and for all $\alpha \in \mathbb{C}$ $\|\alpha v\| = |\alpha|\,\|v\|$.
3. **Triangle Inequality** : For all $v, w \in V$, $\|v + w\| \leq \|v\| + \|w\|$.

Definition 3.2.3 (Normed Vector Space)**.** Let V be a vector space, with some norm $\|\cdot\|$. Then V is called a **Normed Vector Space.** We denote it as $\big(V, \|\cdot\|\big)$.

Theorem 3.2.1. *Let $\big(V, \langle , \rangle\big)$ be an inner product space. Then V is a normed vector space.*

Proof. Define $\|\cdot\| : V \to [\,0,\infty)$ by $v \longmapsto \|v\| := \sqrt{\langle v, v\rangle}$.
We show homogeneity and triangle inequality.

Homogeneity: Take any $\alpha \in \mathbb{C}$ and $v \in V$. Then $\|\alpha v\| = \sqrt{\langle \alpha v, \alpha v\rangle} = \sqrt{\alpha\,\langle \alpha v, v\rangle} = \sqrt{\alpha\overline{\alpha}\,\langle v, v\rangle} = \sqrt{|\alpha|^2\,\langle v, v\rangle} = \sqrt{|\alpha|^2}\sqrt{\langle v, v\rangle} = |\alpha|\,\|v\|$.

Triangle Inequality: Take any two vectors $u, v \in V$ and consider

$$\begin{aligned} 0 \leq \|u+v\|^2 = \langle u+v, u+v\rangle = \langle u, u\rangle + \langle u, v\rangle + \overline{\langle v, u\rangle} + \langle v, v\rangle &= \langle u, u\rangle + 2\,\mathrm{Re}(\langle u, v\rangle) + \langle v, v\rangle \\ &\qquad \text{(by properties of inner product)} \\ &= \langle u, u\rangle + \langle v, v\rangle + 2\,\mathrm{Re}(\langle u, v\rangle) \quad \text{(since complex addition commutes)} \\ &\leq \langle u, u\rangle + \langle v, v\rangle + 2\,|\langle u, v\rangle| \quad \text{((1) see remark below)} \\ &\leq \langle u, u\rangle + 2\,\|u\|\,\|v\| + \langle v, v\rangle \quad \text{(By Cauchy Schwarz Inequality)} \\ &= \big(\,\|u\| + \|v\|\,\big)^2 \end{aligned}$$

Taking the square root of both sides yields the triangle inequality property. ∎

Remark 3.2.5. To see (1) note that $\langle u, v\rangle \in \mathbb{C}$ so it is of the form $a + ib$ for some $a, b \in \mathbb{R}$. Hence, $2\,\mathrm{Re}(\langle u, v\rangle) = 2a \leq 2\,|\langle u, v\rangle| = 2\sqrt{a^2 + b^2}$

Exercise 3.2.5. Prove **Positive-Definiteness** of **Theorem 3.2.1**.

Theorem 3.2.2 (Cauchy Schwarz Inequality)**.** *Let V be an inner product space. Let u, v be any two arbitrary vectors in V. Then*

$$|\langle u, v\rangle| \leq \|u\| \, \|v\|$$

Proof. Let us consider a function f of $t \in \mathbb{R}$, defined by $f(t) := \langle u + tv, u + tv\rangle = \langle u, u\rangle + 2\,\mathrm{Re}(\langle u, tv\rangle) + t\bar{t}\,\langle v, v\rangle = \langle u, u\rangle + 2\,\mathrm{Re}(\langle u, tv\rangle) + t^2\,\langle v, v\rangle$ since t is a real number. Now consider the function $g(t) = \langle u, u\rangle + 2t\,|\langle u, v\rangle| + t^2\,\langle v, v\rangle \geq f(t)$
Note that $g(t)$ is a quadratic function and in general a quadratic function has at most 2 roots. However, $g(t) \geq 0$ since it is defined using the norm function. Hence, g has at most 1 root! That is equivalent to the statement that the discriminant $b^2 - 4ac \leq 0 \iff 4(|\langle u, v\rangle|)^2 - 4\,\langle u, u\rangle\,\langle v, v\rangle \leq 0 \iff |\langle u, v\rangle|^2 \leq \langle u, u\rangle\,\langle v, v\rangle$
Taking the square root of both sides we obtain $|\langle u, v\rangle| \leq \sqrt{\langle u, u\rangle\,\langle v, v\rangle} = \sqrt{\langle u, u\rangle}\sqrt{\langle v, v\rangle} = \|u\|\,\|v\|$ ∎

Exercise 3.2.6. Could we ever obtain equality in **Theorem 3.2.2**?

Now we can define the notion angle between two vectors in a vector space.

Definition 3.2.4. Let V be an inner product space. Let u, v be any two arbitrary vectors in V. Then the **angle between u and v**, θ, is defined to be:

$$\theta = \cos^{-1}\Big(\frac{|\langle u, v\rangle|}{\|u\|\,\|v\|}\Big)$$

And $\theta \in [0, \pi]$

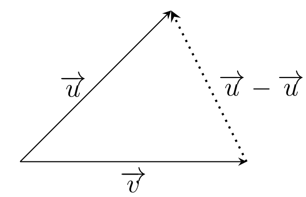
Remark 3.2.6. Recall: $\cos^{-1}(x) = \arccos(x)$!

MOTIVATION: For convenience and simplicity of our arguments suppose we are working with $\mathbb{R}^n$ and more specifically in $\mathbb{R}^2$. In addition, take the usual "dot product" as the inner product defined on $\mathbb{R}^2$. Let $\underline{u} = (u_1, u_2)$ and $\underline{v} = (v_1, v_2)$. By the cosine law we have that

$$\begin{aligned}
& \|\underline{u} - \underline{v}\|^2 = \|\underline{u}\|^2 + \|\underline{v}\|^2 - 2\,\|\underline{u}\|\,\|\underline{v}\|\cos\theta \\
\iff & \langle \underline{u} - \underline{v}, \underline{u} - \underline{v}\rangle = \langle \underline{u}, \underline{u}\rangle + \langle \underline{v}, \underline{v}\rangle - 2\sqrt{\langle \underline{u}, \underline{u}\rangle}\sqrt{\langle \underline{v}, \underline{v}\rangle}\cos\theta \\
\iff & \langle \underline{u}, \underline{u}\rangle - 2\,\langle \underline{u}, \underline{v}\rangle + \langle \underline{v}, \underline{v}\rangle = \langle \underline{u}, \underline{u}\rangle + \langle \underline{v}, \underline{v}\rangle - 2\sqrt{\langle \underline{u}, \underline{u}\rangle\,\langle \underline{v}, \underline{v}\rangle}\cos\theta \\
\iff & \langle \underline{u}, \underline{v}\rangle = \sqrt{\langle \underline{u}, \underline{u}\rangle\,\langle \underline{v}, \underline{v}\rangle}\cos\theta \\
\iff & \langle \underline{u}, \underline{v}\rangle = \|\underline{u}\|\,\|\underline{v}\|\cos\theta \qquad \text{(since square root is multiplicative)}
\end{aligned}$$

Note that in $\mathbb{R}^2$ (or $\mathbb{R}^n$ in general) this inner product(dot product) induces the norm that coincides exactly with our usual perception of the Euclidean length of a vector, i.e.

$$\|\underline{v}\| = \sqrt{\langle \underline{v}, \underline{v}\rangle} = \sqrt{v_1^2 + v_2^2}$$



Definition 3.2.5 (Orthogonal)**.** Let V be an inner product space. Two vectors $u, v \in V$ are said to be **orthogonal** if

$$\langle u, v \rangle = 0$$

Remark 3.2.7. If V is an inner product space and S is any subset of V then $S^{\perp} = \{u : \langle u, v \rangle = 0 \; \forall \; v \in S\}$ would be a subspace of V. We are going to study this important subspace in more detail very soon.

Theorem 3.2.3. *Let $(V, \langle , \rangle)$ be an inner product space. Let u, v be any two orthogonal vectors in V. Then $\|u + v\|^2 = \|u\|^2 + \|v\|^2$.*

Exercise 3.2.7. Prove **Theorem 3.2.3**, which is simply the generalized version of **Pythagorean Theorem.**

Remark 3.2.8. Whenever we have an inner product, $\|.\|$ is automatically perceived to be the **induced norm.**

Now that we have seen how the notion of a norm can be induced by an inner product space, we could alternatively define a notion of "distance." However, let us first define an independent function that can capture the distance between two vectors regardless of whether or not we have an inner product or even norm.

Definition 3.2.6 (Metric)**.** Let V be a vector space. A **metric** d is a function defined on $V \times V$ into $[0, \infty)$ with following properties:

- $\forall u, v \in V, d((u, v)) = 0$ if and only if $u = v$.
- **Symmetry:** $\forall u, v \in V$, $d((v, u)) = d((u, v))$
- **Triangle inequality:** $\forall u, v, w \in V$, $d((u, w)) \leq d((u, v)) + d((v, w))$.

Example 3.2.4. The discrete metric d on a vector space V is defined as follows:

$$\forall u, v \in V, \; d((u, v)) = \begin{cases} 0 & \text{if } u = v, \\ 1 & \text{otherwise.} \end{cases}$$

Exercise 3.2.8. Verify that d in the above example is a metric on V.

Theorem 3.2.4. *Given any normed vector space $(V, \|\cdot\|)$, $d : V \times V \longmapsto [0, \infty)$ defined by $d((u, v)) = \|u - v\|$ is a metric on V. We say that d is a norm-induced metric.*

Proof. We verify by definition that d is a metric on V. Let $u, v, w \in V$.

- $d((u, v)) = \|u - v\| \geq 0$ by positivity of $\|\cdot\|$.
  Also $d((u, v)) = \|u - v\| = 0 \iff u - v = 0 \iff u = v$ by definiteness of $\|\cdot\|$.
- Now we check symmetry.

$$\begin{aligned} d((u, v)) &= \|u - v\| \\ &= \|-(v - u)\| && \text{(Additive inverse of } u - v\text{)} \\ &= \|(-1)(v - u)\| && \text{(Theorem 2.1.2)} \\ &= |-1| \, \|(v - u)\| && \text{(homogeneity of norm)} \\ &= \|v - u\| = d((v, u)) \end{aligned}$$

- Triangle inequality:

$$\begin{aligned} d((u, w)) &= \|u - w\| \\ &= \|(u - v) + (v - w)\| \\ &\leq \|u - v\| + \|v - w\| && \text{(triangle inequality for } \|\cdot\|\text{)} \\ &= d((u, v)) + d((v, w)) \end{aligned}$$

■

Example 3.2.5. Compute the norm of the vector $p(x) = x + 1$ in the inner product space $\mathcal{C}([0, 1], \mathbb{R})$ with the inner product defined, and $d((x^2, x))$.
SOL: We know that for all $f(x), g(x) \in \mathcal{C}([0, 1], \mathbb{R})$, we have that

$$\langle f(x), g(x) \rangle = \int_0^1 f(x)g(x)dx$$

So

$$\begin{aligned} \|x + 1\| &= \sqrt{\int_0^1 (x + 1)^2 dx} \\ &= \sqrt{\int_0^1 (x^2 + 2x + 1) dx} \\ &= \sqrt{\frac{x^3}{3} + x^2 + x\Big|_0^1} \\ &= \sqrt{\frac{7}{3}} \end{aligned}$$

We can also see that

$$\begin{aligned} d((x^2, x)) &= \|x^2 - x\| \\ &= \sqrt{\int_0^1 (x^2 - x)(x^2 - x)dx} \\ &= \sqrt{\int_0^1 x^4 - 2x^3 + x^2 dx} \\ &= \sqrt{\frac{x^5}{5} - \frac{x^4}{2} + \frac{x^3}{3}\Big|_0^1} \\ &= \frac{1}{\sqrt{30}} \end{aligned}$$

Example 3.2.6. Consider the vector space $\mathcal{C}([0,1], \mathbb{R})$ over $\mathbb{R}$ again. In Example 3.2.5 we worked with the **norm-induced metric** d on this space However, let us define another **metric** on this vector space without using any inner product function:

$$\begin{aligned} &d : \mathcal{C} \times \mathcal{C} \to [0, \infty) \\ &d(f(x), g(x)) \longmapsto \int_0^1 |f(x) - g(x)|\, \psi(x) dx \quad \text{(where } \psi : [0,1] \to (0, \infty) \text{ is any fixed continuous function)} \end{aligned}$$

Exercise 3.2.9. Prove the function d defined in Example 3.2.6 is indeed a metric on $\mathcal{C}([0,1], \mathbb{R})$.

Remark 3.2.9. Let us investigate a **third metric**, measuring distances between the functions in $\mathcal{C}([0,1], \mathbb{R})$. We can define $d((f(x), g(x))$ to be the maximum vertical distance between their graphs of $f(x)$ and $g(x)$ over the region [0,1]. Let us see an example. Consider the functions $f(x) = x^2$ and $g(x) = x$. We can use calculus to find that distance by differentiating $x^2 - x$ and finding its critical points, showing its maximum occurs at $x = \frac{1}{2}$ with a value of $\frac{1}{4}$. Note that the distance induced by the inner product (the integral function) yields a much smaller distance, $\frac{1}{\sqrt{30}}$ in contrast to the distance $\frac{1}{4}$ obtained from the maximum distance between their graphs.

The following notion allows us to investigate another important subspace.

Definition 3.2.7 (Orthogonal Complement)**.** Let V be an inner product space and W be any subspace of V. Then the **orthogonal complement** $W^\perp = \{v \in V \mid \langle v, w\rangle = 0\ \forall w \in W\}$ is the set of all vectors that are orthogonal to any vector in W.

Remark 3.2.10. $W^\perp$ is a subspace, since if $z_1, z_2 \in W^\perp$ and $\alpha \in \mathbb{C}$ and $u \in W$, then $\langle u, z_1 + \alpha z_2\rangle = \langle u, z_1\rangle + \alpha\langle u, z_2\rangle = 0$.

This particular subspace allows us to decompose our space (must be finite dimensional) into two parts as follows:

Theorem 3.2.5. *Let V be a finite dimensional inner product space and W is any subspace of V. Then $V = W \oplus W^{\perp}$.*

We will prove this theorem soon, but first let us introduce another notion that will be directly related to computing the orthogonal complement of a given subspace. That is the notion of projection.

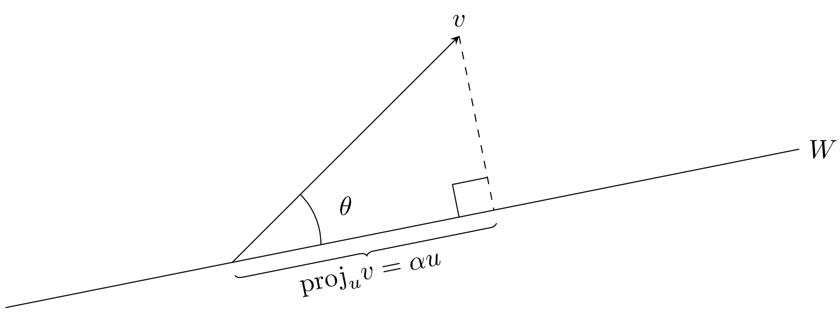
Motivation: Let $(V, \langle , \rangle)$ be a finite dimensional vector space and W be any subspace of V. Since we have a notion of a metric (distance) on this vector space, we can find the closest distance from an arbitrary vector v to the subspace W. Clearly, if $v \in W$, the answer is simply 0. How about if $v \notin W$? Now it's the notion of "projection of v onto W" that comes into play.

For simplicity and convenience of our argument, let us first restrict ourselves to $V = \mathbb{R}^2$ and W be any 1-dimensional subspace of V, technically a line in the plane.

Consider an arbitrary vector $\underline{v} = (v_1, v_2) \in \mathbb{R}^2$ and $W = \text{sp}(\underline{u})$ for some $\underline{u} = (u_1, u_2) \in \mathbb{R}^2$, where

$$W = \{\alpha \underline{u} : \alpha \in \mathbb{R}\}$$

Since the projection of $\underline{v}$ onto W is an element in the subspace W, it must be of the form $\alpha \underline{u}$ for some $\alpha \in \mathbb{R}$



We know that

$$\cos\theta = \frac{\|\alpha u\|}{\|v\|} = \frac{|\alpha|\|u\|}{\|v\|} \Rightarrow |\alpha| = \cos\theta \frac{\|v\|}{\|u\|} \tag{1}$$

and

$$\langle u, v\rangle = \cos\theta \|u\|\|v\| \Rightarrow \cos\theta = \frac{\langle u, v\rangle}{\|u\|\|v\|} \tag{$*$}$$

Hence, substituting $(*)$ into (1), we obtain that

$$\alpha = \frac{\langle u, v\rangle}{\|u\|^2}. \quad \therefore \text{proj}_u v = \frac{\langle u, v\rangle}{\|u\|^2} u$$

Note that $\langle \alpha u, v - \alpha u\rangle = 0$ since

$$\begin{aligned}\langle \alpha u, v - \alpha u\rangle &= \overline{\alpha}\langle u, v - \alpha u\rangle \\ &= \overline{\alpha}(\langle u, v\rangle - \alpha\langle u, u\rangle) \\ &= \alpha\left(\langle u, v\rangle - \frac{\langle u, v\rangle}{\|u\|^2}\langle u, u\rangle\right) \\ &= \alpha\left(\langle u, v\rangle - \langle u, v\rangle\right) \\ &= 0\end{aligned}$$

Now, how do we project v onto a subspace W where $\dim(W) \geq 2$? Note that W must be of finite dimension for the idea of projection to make mathematical sense. But V does not need to be finite! On any inner product space V, as long as W is a **finite** subspace of V we have that proj_W is a linear map from V to V. You will investigate the properties of projection map on *assignment* #3.

Now let us first suppose that V is **finite dimensional** inner product space. Therefore, given any subspace W of V we know that $V = W \oplus W^{\perp}$. (We will prove this on page 57) Since V is a direct sum of W and its orthogonal compliment, we can fix a basis $A = \{w_1, w_2, \ldots, w_k\}$ for W and a basis $B = \{w_{k+1}, \ldots, w_n\}$ for $W^{\perp}$, and we know $\{w_1, \cdots, w_n\}$ is a basis of V. (We only need to show it is linearly independent since by definition W and $W^{\perp}$ must span V. To this end if $\sum_{i=1}^{n} \gamma_i w_i = 0_V$ then we have $\sum_{i=1}^{k} \gamma_i w_i = -\sum_{i=k+1}^{n} \gamma_i w_i$ and since $W \cap W^{\perp} = \{0_V\}$ independence of each of those two sets A, B independently implies that $\gamma_i = 0 \;\forall\; 1 \leq i \leq n$)

Fix a $v \in V$, and let us focus on what we want to obtain; finding the $\text{proj}_W v$. Since $A \cup B$ is a basis of V, there exists scalars α_i such that $v = \sum_{i=1}^{n} \alpha_i w_i$, and this linear combination is unique. Now note that $\text{proj}_W v$ yields a vector in W which is closest in distance to v and such vector is unique, hence, if we decompose the linear combination $\sum_{i=1}^{n} \alpha_i w_i$ to $\sum_{i=1}^{k} \alpha_i w_i + \sum_{i=k+1}^{n} \alpha_i w_i$, then

Remark 3.2.11. $\text{proj}_W v = \sum_{i=1}^{k} \alpha_i w_i$ and $\sum_{i=k+1}^{n} \alpha_i w_i$ would represent $\text{proj}_{W^{\perp}} v$. The latter part is only true if V is finite dimensional.

Therefore, in order to find the projection of v onto W, we only need to find the scalars $\alpha_1, \cdots, \alpha_k$. To this end, pick any basis element of W, say w_i for any i, where $1 \leq i \leq k$. For simplicity let us take $i = 1$. Now $\langle w_1, v\rangle$ then we have that:

$$\left\langle w_1, \sum_{i=1}^{k} \alpha_i w_i + \sum_{i=k+1}^{n} \alpha_i w_i \right\rangle = \sum_{i=1}^{k} \alpha_i \langle w_1, w_i\rangle + \sum_{i=k+1}^{n} \alpha_i \langle w_1, w_i\rangle$$

Since $w_i \in W^{\perp}$ for all $k+1 \leq i \leq n$, we have $\langle w_1, w_i\rangle = 0$.

Hence, $\langle w_1, v\rangle$ reduces down to $\sum_{i=1}^{k} \alpha_i \langle w_1, w_i\rangle = \alpha_1\langle w_1, w_1\rangle + \cdots + \alpha_k\langle w_1, w_k\rangle$.

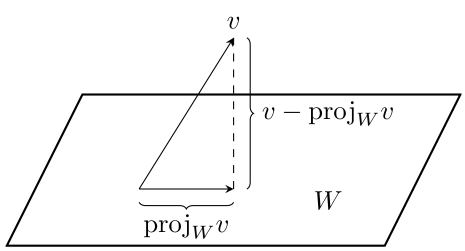
Note that $\langle w_1, w_i\rangle \in \mathbb{F}$, where $\mathbb{F}$ is either $\mathbb{R}$ or $\mathbb{C}$, for any $1 \le i \le k$. Also, each of these scalars is associated to the unknown variable α_i, which we are aiming to find. So we have one equation with k many unknowns. But to find these unique α_i we need to have k many equations. Note we can easily construct them by considering $\langle w_i, v\rangle$ for $1 < i \le k$, where each of them reduces down to $\sum_{i=1}^{k} \alpha_i \langle w_1, w_i\rangle$. Then we can consider the homogeneous linear system of equations, form the matrix of coefficients and apply elementary row operation to reduce the matrix to find the values of α_i for all $1 \le i \le k$. But this process can be very cumbersome. However, we could have simplified this process had we imposed the condition that the basis $\{w_1, \ldots, w_k\}$ elements are **mutually orthogonal**, that is, $\langle w_j, w_i\rangle = 0$ for all $i \ne j$, then $\langle w_i, v\rangle$ can even be reduced further down to $\langle w_i, \alpha_i w_i\rangle$. So, $\langle w_i, v\rangle = \alpha_i \langle w_i, w_i\rangle = \alpha_i \|w_i\|^2$. Hence, we can easily find that $\alpha_i = \frac{\langle w_i, v\rangle}{\|w_i\|^2}$

Thereby we can formulate the following proposition.

Proposition 3.2.2. *Let V be a finite inner product space and W be a subspace. Suppose $\{w_1, \ldots, w_k\}$ be an orthogonal basis of W. Then the projection of v onto W, $\mathrm{proj}_W v$, is defined to be*

$$\mathrm{proj}_W v = \frac{\langle w_1, v\rangle}{\|w_1\|^2} w_1 + \frac{\langle w_2, v\rangle}{\|w_2\|^2} w_2 + \cdots + \frac{\langle w_k, v\rangle}{\|w_k\|^2} w_k$$

Remark 3.2.12. $\mathrm{proj}_W v$ is the vector in W which is closest to the vector v, with distance $\|v - \mathrm{proj}_W v\|$. Recall that distance $d((v, \mathrm{proj}_W v)) = \|v - \mathrm{proj}_W v\|$. Equivalently $v - \mathrm{proj}_W\, v$ is orthogonal to every vector in W.



Example 3.2.7. Find the projection of the vector $(3, 1, -2)$ onto the plane $W = \{(x, y, z) \mid x + y + z = 0\}$.

Answer: We could find an orthogonal basis of the subspace W, which is a 2-dimensional subspace of $\mathbb{R}^3$, then project v onto these two orthogonal basis elements. However, there is an easier approach.

Note $W^\perp = \mathrm{sp}((1, 1, 1))$. Any vector on the plane will be orthogonal to any $\alpha(1, 1, 1)$ for any $\alpha \in \mathbb{R}$.
But note $V = W \oplus W^\perp$ so by remark 3.2.11 $v = \mathrm{proj}_W v + \mathrm{proj}_{W^\perp} v$.

$$\therefore v - \mathrm{proj}_{W^\perp} v = \mathrm{proj}_W v$$

We can easily find $\mathrm{proj}_{W^\perp} v = \dfrac{\langle (1, 1, 1), v\rangle}{\|(1, 1, 1)\|^2}(1, 1, 1)$.

Recall we are using the usual inner product on $\mathbb{R}^3$!

$$\therefore \operatorname{proj}_{W^\perp} v = \frac{\langle (1,1,1),(3,1,-2)\rangle}{\|(1,1,1)\|^2}(1,1,1) = \frac{3+1-2}{3}(1,1,1) = \frac{2}{3}(1,1,1)$$

$$\therefore \operatorname{proj}_W v = (3,1,-2) - \left(\frac{2}{3}(1,1,1)\right) = (3,1,-2) - \frac{2}{3}(1,1,1)$$

Lemma 3.2.1. *Given any set $\{w_1,\dots,w_n\}$ which is mutually orthogonal of non-zero vectors in an inner product space, $\{w_1,\dots,w_n\}$ is linearly independent.*

Proof. Suppose $\{w_1,\dots,w_n\}$ is linearly dependent.

$$\Rightarrow \exists \alpha_j \neq 0 \text{ such that } \sum_{i=1}^n \alpha_i w_i = 0$$

WLOG, let $\alpha_j = \alpha_1 \neq 0$,

$$\Rightarrow \alpha_1 w_1 + \sum_{i=2}^n \alpha_i w_i = 0$$

$$\Rightarrow \alpha_1 w_1 = -\sum_{i=2}^n \alpha_i w_i$$

$$\Rightarrow w_1 = \alpha_1^{-1}\left(-\sum_{i=2}^n \alpha_i w_i\right) = -\alpha_1^{-1}\sum_{i=2}^n \alpha_i w_i$$

Now consider

$$\begin{aligned} \|w_1\|^2 = \langle w_1, w_1\rangle &= \left\langle w_1, -\alpha_1^{-1}\sum_{i=2}^n \alpha_i w_i\right\rangle \\ &= -\alpha_1^{-1}\sum_{i=2}^n \alpha_i \langle w_1, w_i\rangle && \text{(by linearity in the } 2^{nd} \text{ slot)} \\ &= 0 && \text{(since } \{w_1,\dots,w_n\} \text{ are pairwise orthogonal)} \end{aligned}$$

But $w_1 \neq 0_V$, but that's a contradiction to the positive definiteness condition of norm. ∎

The following process will allow us to find an orthogonal basis given any basis.

Lemma 3.2.2 (Gram-Schmidt Process)**.** *Let V be any inner product space and W be any finite dimensional subspace of V. Then W has an orthogonal basis, in particular, an orthonormal basis.*

Proof. Let $\{v_1, v_2, \ldots, v_n\}$ be a basis of W.
We proceed to define an orthogonal basis $\{b_1, b_2, \ldots, b_n\}$ of W as follows:
Let $b_1 = v_1$.
To find b_2, project v_2 onto b_1 and subtract it from v_2. So,

$$b_2 = v_2 - \operatorname{proj}_{\operatorname{sp}(b_1)} v_2 = v_2 - \frac{\langle b_1, v_2 \rangle}{\|b_1\|^2} b_1$$

To be more specific, we project v_2 onto the subspace generated by b_1, $\operatorname{sp}(b_1)$. So clearly,

$$\langle b_1, b_2 \rangle = 0$$

Hence, to find b_i, we project v_i onto the subspace spanned by $\{b_1, \ldots, b_{i-1}\}$ (which are already mutually orthogonal and subtract it from v_i to find the orthogonal vector to every vector $v_1, \ldots, v_{i-1}$.

$$b_i = v_i - \operatorname{proj}_{\operatorname{sp}(b_1,\ldots,b_{i-1})} v_i = v_i - \left(\frac{\langle b_1, v_i \rangle}{\|b_1\|^2} b_1 + \frac{\langle b_2, v_i \rangle}{\|b_2\|^2} b_2 + \cdots + \frac{\langle b_{i-1}, v_i \rangle}{\|b_{i-1}\|^2} b_{i-1} \right)$$

by Proposition 3.2.2.
Now to find an orthonormal basis, note that given any $v \in V$, the vector $q = \frac{1}{\|v\|} v$ is normalized. That is,

$$\|q\| = \left\| \frac{1}{\|v\|} v \right\| = \frac{1}{\|v\|} \|v\| = 1$$

So we can normalize the orthogonal basis $\{b_1, \ldots, b_n\}$ as follows:

$$\{q_1 = \frac{b_1}{\|b_1\|}, q_2 = \frac{b_2}{\|b_2\|}, \ldots, q_n = \frac{b_n}{\|b_n\|}\}$$

∎

Remark 3.2.13. Hence, in Proposition 3.2.2, if $\{w_1, \ldots, w_k\}$ is an orthonormal basis, then $\operatorname{proj}_W v = \langle w_1, v \rangle w_1 + \ldots + \langle w_k, v \rangle w_k$.

Theorem 3.2.6. *Let V be a finite inner product space of dimension n, and $A = \{b_1, \ldots, b_i\}$ be any orthogonal set of vectors. Then A can be extended to an orthogonal basis for V. In particular, any finite inner product space has an orthonormal basis.*

Proof. Apply the Gram-Schmidt process starting off with v_{i+1}.
First, we know that $\{b_1, b_2, \ldots, b_i\}$ is a linearly independent set and hence we can extend to a basis of V:

$$\{b_1, b_2, \ldots, b_i, v_{i+1}, \ldots, v_n\}$$

and generate b_{i+1} by projecting v_{i+1} on the span of $\{b_1, b_2, \ldots, b_i\}$ and subtracting it from v_{i+1}.
Clearly, we can easily normalize any set of non-zero vectors. ∎

Remark 3.2.14. Let V be a finite inner product space and $\{q_1, q_2, \ldots, q_n\}$ be an orthonormal basis of V. Then given $v \in V$, we know

$$v = \alpha_1 q_1 + \cdots + \alpha_n q_n$$

for $\alpha_i \in \mathbb{R}$.
Hence, note we can express

$$v = \langle q_1, v\rangle q_1 + \langle q_2, v\rangle q_2 + \cdots + \langle q_n, v\rangle q_n$$

and hence

$$\|v\|^2 = |\langle q_1, v\rangle|^2 + |\langle q_2, v\rangle|^2 + \cdots + |\langle q_n, v\rangle|^2 \qquad (*)$$

Note the equality $(*)$ is simply the generalized Pythagorean theorem.

Now we can prove a stronger version of 3.2.5.

Theorem 3.2.7. *Let V be any inner product space (Note: in 3.2.5, V was finite). Let W be any finite dimensional subspace. Then $V = W \oplus W^{\perp}$.*

Proof. By Lemma 3.2.2, W yields an orthonormal basis. Let $\{q_1, q_2, \ldots, q_n\}$ be an orthonormal basis of W. For any $v \in V$, we only need to establish that v can be expressed as a unique linear combination of an element of W with an element in $W^{\perp}$.
Note: $\forall\ 1 \leq i \leq n$ then $\langle q_i, v\rangle \in \mathbb{C}$, and since $q_i \in W$ and W is a subspace of V,

$$\sum_{i=1}^{n} \langle q_i, v\rangle q_i \in W$$

Now,

$$v = \sum_{i=1}^{n} \langle q_i, v\rangle q_i + \underbrace{\left(v - \sum_{i=1}^{n} \langle q_i, v\rangle q_i\right)}_{w}$$

Clearly, we have $w \in W^{\perp}$ as we have shown previously, and denoted as

$$\sum_{i=1}^{n} \langle q_i, v\rangle q_i = \mathrm{proj}_W v \quad \text{by proposition 3.2.2}$$

■

Corollary 3.2.7.1. *Let V be any inner product space and W be a finite dimensional subspace of V. Then $(W^{\perp})^{\perp} = W$. Furthermore, if V is finite dimensional then $\dim W^{\perp} = \dim V - \dim W$.*

Proof. Let $v \in W$. Then if $w \in W^{\perp}$ then $\langle v, w\rangle = 0$ by definition of orthogonal complement. So, for every $w \in W^{\perp}$, $\langle v, w\rangle = 0$ implies that v is an element of the orthogonal complement of $W^{\perp}$. In particular, $v \in (W^{\perp})^{\perp}$. Hence in general we always have that $W \subseteq (W^{\perp})^{\perp}$ for any subspace!

Now let us show $(W^{\perp})^{\perp} \subseteq W$. Pick $v \in (W^{\perp})^{\perp}$. Since $(W^{\perp})^{\perp} \subseteq V$ and $V = W \oplus W^{\perp}$ by Theorem 3.2.5, there exists unique $u \in W$ and $z \in W^{\perp}$ such that $v = u + z$.
Note that $v - z \in W$ since $v - z = u$. But we always have $W \subseteq (W^{\perp})^{\perp}$ so $u = v - z \in (W^{\perp})^{\perp}$ as well. Now note that the orthogonal complement of any subset is a subspace. Hence, since $v \in (W^{\perp})^{\perp}$ and $u \in (W^{\perp})^{\perp}$, we get $v - u \in (W^{\perp})^{\perp}$, but $z = v - u$ so $z \in (W^{\perp})^{\perp}$ and $z \in W^{\perp}$. It follows that $0 = \langle z, z\rangle = \langle v - u, v - u\rangle$. Now by properties of $\langle , \rangle$, we must have $v - u = 0 \iff v = u \in W$.
Therefore $(W^{\perp})^{\perp} \subseteq W$. ■

Example 3.2.8. Find an orthonormal basis for the subspace $\mathrm{sp}(1, \sqrt{x}, x)$ of the inner product space $\mathcal{C}([0,1], \mathbb{R})$.

Answer: Let $b_1 = 1$. Then

$$\text{proj}_{\text{sp}(1)}\sqrt{x} = \frac{\langle 1, \sqrt{x}\rangle}{\|1\|^2}1 = \frac{\langle 1, \sqrt{x}\rangle}{\langle 1, 1\rangle}1 = \frac{\int_0^1 \sqrt{x}\,dx}{\int_0^1 1\,dx}1 = \frac{\frac{2}{3}}{1}1 = \frac{2}{3}$$

Hence, $b_2 = \sqrt{x} - \frac{2}{3}$.

Now let $E = \text{sp}(1, \sqrt{x} - \frac{2}{3})$, then

$$\begin{aligned} x - (\text{proj}_E x) &= x - \left(\frac{\langle 1, x\rangle}{\|1\|^2}1 + \frac{\langle \sqrt{x} - \frac{2}{3}, x\rangle}{\left\|\sqrt{x} - \frac{2}{3}\right\|^2}\left(\sqrt{x} - \frac{2}{3}\right)\right) \\ &= x - \left(\frac{\langle 1, x\rangle}{\langle 1, 1\rangle}1 + \frac{\langle \sqrt{x} - \frac{2}{3}, x\rangle}{\langle \sqrt{x} - \frac{2}{3}, \sqrt{x} - \frac{2}{3}\rangle}\left(\sqrt{x} - \frac{2}{3}\right)\right) \\ &= x - \frac{\int_0^1 x\,dx}{\int_0^1 1\,dx}1 - \frac{\int_0^1 x(\sqrt{x} - \frac{2}{3})\,dx}{\int_0^1 (\sqrt{x} - \frac{2}{3})^2\,dx}\left(\sqrt{x} - \frac{2}{3}\right) \\ &= x - \frac{1}{2}\cdot 1 - \frac{6}{5}(\sqrt{x} - \frac{2}{3}) \\ &= x - \frac{6\sqrt{x}}{5} + \frac{3}{10} \end{aligned}$$

Now to make computation easier, we scale $x - (\text{proj}_E x)$ by 10 and let $b_3 = 10(x - (\text{proj}_E x)) = 10x - 12\sqrt{x} + 3$. Note that scaling by a nonzero scalar doesn't change the span or the orthogonality.
Finally, we normalize each one of b_1, b_2, b_3 to get an orthonormal basis:

$$\frac{b_1}{\|b_1\|} = \frac{1}{\|1\|} = \frac{1}{\sqrt{\langle 1, 1\rangle}} = \frac{1}{\sqrt{\int_0^1 1\,dx}} = 1$$

$$\frac{b_2}{\|b_2\|} = \frac{\sqrt{x} - \frac{2}{3}}{\left\|\sqrt{x} - \frac{2}{3}\right\|} = \frac{\sqrt{x} - \frac{2}{3}}{\sqrt{\langle \sqrt{x} - \frac{2}{3}, \sqrt{x} - \frac{2}{3}\rangle}} = \frac{\sqrt{x} - \frac{2}{3}}{\sqrt{\int_0^1 (\sqrt{x} - \frac{2}{3})^2\,dx}} = \sqrt{2}(3\sqrt{x} - 2)$$

Similarly,

$$\frac{b_3}{\|b_3\|} = \frac{10x - 12\sqrt{x} + 3}{\|10x - 12\sqrt{x} + 3\|} = \frac{10x - 12\sqrt{x} + 3}{\sqrt{\int_0^1 (10x - 12\sqrt{x} + 3)^2\,dx}} = \sqrt{3}\left(10x - 12\sqrt{x} + 3\right)$$

Therefore, an orthonormal basis would be $\left\{1, \sqrt{2}(3\sqrt{x} - 2), \sqrt{3}\left(10x - 12\sqrt{x} + 3\right)\right\}$.

3.3 Projection Matrix

Let V be an inner product space and W be a finite dimensional subspace. Then we have shown that $\text{proj}_W : V \to V$ is a linear map with $\text{range}(\text{proj}_W) = W$ and $\ker(\text{proj}_W) = W^\perp$. Now if V is finite dimensional, using the fact that every linear map yields a matrix representation we are going to proceed to define the matrix representation of the projection map. This will definitely facilitate our computations. We no longer need to find an orthogonal basis of W and apply the projection formula.

Definition 3.3.1. We call the matrix representation of proj_W the **projection matrix** P.

Remark 3.3.1. If we equip the set $\mathcal{L}(V)$ with pointwise addition of linear maps $+$ and composition of linear maps $\circ$, then $(\mathcal{L}(V), +, \circ)$ is an algebraic structure. Note $\circ$ is **not** commutative, however, it does satisfy the associativity and distributivity over addition. Also, not every linear map is invertible; hence we don't have multiplicative inverses, but we do have a multiplicative identity, the identity map $I(v) = v \ \forall v \in V$.
Now, $\text{proj}_W \circ \text{proj}_W = \text{proj}_W$. Hence, it is an idempotent element of $\mathcal{L}(V)$.

Now let us investigate how to construct the projection matrix P. Let $V = \mathbb{R}^n$ and $W \subseteq V$ of dimension m.

$$\begin{aligned} \text{proj}_W &: \mathbb{R}^n \to \mathbb{R}^n \\ &\underline{x} = (x_1, x_2, \ldots, x_n) \longmapsto \text{proj}_W \underline{x} \in W \end{aligned}$$

First, fix any basis $\{\underline{v_1}, \ldots, \underline{v_m}\}$ of W. Form the matrix $A = \begin{bmatrix} | & & | \\ \underline{v_1} & \cdots & \underline{v_m} \\ | & & | \end{bmatrix}_{n\times m}$. Clearly, $\text{col}(A) = W$ since every element in W can be written as a linear combination of $\{\underline{v_1}, \ldots, \underline{v_m}\}$.
Now pick $\underline{x} \in \mathbb{R}^n$. Then $\underline{x}_W = \text{proj}_W \, \underline{x} \in W$, so there exists unique scalars $\alpha_1, \ldots, \alpha_m \in \mathbb{R}$ such that

$$\underline{x}_W = \sum_{i=1}^{m} \alpha_i \underline{v_i} = \alpha_1(v_{11}, v_{12}, \cdots, v_{1n}) + \cdots + \alpha_m(v_{m1}, v_{m2}, \cdots, v_{mn}).$$

Hence, $\exists$ a vector $\underline{c} = (\alpha_1, \ldots, \alpha_m)$ such that

$$Ac^T = \begin{bmatrix} v_{11} & v_{21} & \cdots & v_{m1} \\ v_{12} & v_{22} & \cdots & v_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ v_{1n} & v_{2n} & \cdots & v_{mn} \end{bmatrix}_{n\times m} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_m \end{bmatrix}_{m\times 1} = x^T_{W\,n\times 1}.$$

Note $\underline{c}$ is a vector and c^T is the corresponding column matrix of the vector $\underline{c}$ of size $m \times 1$.

But note $\underline{x} = \underbrace{\underline{x}_W}_{\in W} + \underbrace{\underline{x} - \underline{x}_W}_{\in W^\perp} \implies x^T - x_W^T = x^T - Ac^T = x^T - x_W^T \in W^\perp. \quad (1)$

Let us view $x^T - Ac^T$ simply as a vector(a column vector) in $\mathbb{R}^n$ as opposed to a column matrix of size $n \times 1$. It is just that if I want to be consistent with my notations and express a vector (row vector) in $\mathbb{R}^n$, $(a_1, \cdots, a_n)$, I would have to write the column matrix $x^T - Ac^T$ as $\underline{x} - (\underline{Ac^T})^T = \underline{x} - \underline{c}A^T$, but it is simply too much! So many symbols need to be utilized. Also, note that I will not care if you simply just write x or $\underline{x}$ on your exam and view it as a vector in $\mathbb{R}^n$ on some instances or a matrix of size $n \times 1$ on other instances. I will automatically assume you know from the context which is the case.

Now if we consider $A^T = \begin{bmatrix} v_{11} & v_{12} & \cdots & v_{1n} \\ v_{21} & v_{22} & \cdots & v_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ v_{m1} & v_{m2} & \cdots & v_{mn} \end{bmatrix}_{m \times n}$ and find its nullspace, then if $(z_1, \ldots, z_n) = \underline{z} \in \text{Null}(A^T)$ then

$$A^T z^T = \begin{bmatrix} v_{11} & v_{12} & \cdots & v_{1n} \\ v_{21} & v_{22} & \cdots & v_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ v_{m1} & v_{m2} & \cdots & v_{mn} \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \iff \forall 1 \leq i \leq m, \ \underline{v}_i \cdot \underline{z} = \langle \underline{v}_i, \underline{z} \rangle = 0$$

Hence, finding a basis for the nullspace of A^T gives us a basis for $W^\perp$, i.e. $\text{col}(A)^\perp = W^\perp = \text{Null}(A^T)$.

Now by equation

$$(1) \quad x^T - Ac^T \in W^\perp = \text{Null}(A^T) \implies A^T(x^T - Ac^T) = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \implies A^T x^T = A^T A c^T.$$

Now, multiplying by the inverse of $A^T A$ (why is $A^T A$ invertible?), we get

$$(A^T A)^{-1} A^T x^T = c^T \implies A(A^T A)^{-1} A^T x^T = Ac^T = x_W^T = \underline{x}_W = \text{proj}_W \underline{x}$$

Therefore, $P = A(A^T A)^{-1} A^T$.

As P is the projection matrix, P is the matrix representation of the projection map, and proj_W is idempotent, so is P, that is $P \cdot P = P^2 = P$. Recall that the set of all square matrices forms an **algebraic structure** under the usual matrix addition and multiplication.

Now, P possesses another algebraic property. P is **symmetric**: $P^T = P$.

Note:

$$\begin{aligned} P^2 &= (A(A^T A)^{-1} A^T)(A(A^T A)^{-1} A^T) \\ &= A(A^T A)^{-1}(A^T A)(A^T A)^{-1} A^T \\ &= A(A^T A)^{-1} A^T = P. \end{aligned}$$

And clearly,

$$\begin{aligned} P^T &= (A(A^T A)^{-1} A^T)^T \\ &= (A^T)^T((A^T A)^{-1})^T A^T \quad \text{(property of transpose)} \\ &= A((A^T A)^T)^{-1} A^T \quad \text{(property of transpose)} \\ &= A(A^T (A^T)^T)^{-1} A^T. \quad \text{(property of transpose)} \\ &= A(A^T A)^{-1} A^T = P \end{aligned}$$

Exercise 3.3.1. The projection matrix is defined assuming that A^TA is invertible. We would be in big trouble if it's not! Luckily, A^TA turns out to be invertible. Can you show it?
Hint: $\text{Null}(A^TA) = \text{Null}(A)$, and A^TA is a square matrix.

Theorem 3.3.1 (Rank of A^TA)**.** *Let A be an $n \times m$ real matrix. Then A^TA is a symmetric matrix and rank$(A^TA) =$ rank(A).*

Proof. By properties of transpose,

$$(A^TA)^T = A^T(A^T)^T = A^TA.$$

Or you can simply apply the usual matrix multiplication.
We show if $\text{Null}(A^TA) = \text{Null}(A) \implies \text{rank}(A) = \text{rank}(A^TA)$ is in the nullspace of A.
Let $v = (v_1, \ldots, v_m)$ be in the nullspace of A.

Hence, $Av^T = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}_{n\times 1}$. Then clearly, $(A^TA)\underline{v}^T = A^T(Av^T) = A^T{}_{m\times n}\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}_{n\times 1} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}_{m\times 1}$.
So v is in the nullspace of A^TA.

Now let $w = (w_1, \ldots, w_m)$ be in the nullspace of the square matrix $(A^TA)_{m\times m}$.
Then

$$A^TAw^T = A^TA\begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_m \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}_{m\times 1}$$

Now multiply both sides of this equation by the row matrix $\begin{bmatrix} w_1 & w_2 & \cdots & w_m \end{bmatrix}$ on the left:

$$\begin{bmatrix} w_1 & w_2 & \cdots & w_m \end{bmatrix} A^TAw^T = [0]_{1\times 1}.$$

So, $wA^T \cdot Aw^T = 0$ if we view them as vectors.
Thereby viewing Aw^T as a vector; in particular a column vector in $\mathbb{R}^n$ rather than an $n \times 1$ column matrix we get the Euclidean inner product (the usual "dot product") :

$$\langle Aw^T, Aw^T \rangle = \|Aw^T\|^2 = 0.$$

By the properties of norm, it must be that $Aw^T = (0, 0, \cdots, 0) \in \mathbb{R}^n$. So w^T is in the nullspace of A as well. ∎

Theorem 3.3.2 (Characterization of Projection Matrices). *The projection matrix for a subspace $W \subseteq \mathbb{R}^n$ is both idempotent and symmetric. Conversely, given any $n \times n$ matrix which is both symmetric and idempotent, it is a projection matrix.*

Proof. We have shown that the projection matrix is both symmetric and idempotent using the formula of the projection matrix. Equivalently, note that the matrix representation of the composition of two linear maps is the product of their matrix representations. We already established that (in Assignment 3) the composition $\text{proj}_W \circ \text{proj}_W = \text{proj}_W$. Hence, $P^2 = P \cdot P = P$.
For the converse direction, let P be an $n \times n$ idempotent and symmetric matrix. To show it's a projection matrix, first we need to determine a subspace onto which we're projecting. Hence, let $W = \text{col}(P)$, the column space of P.
But first, recall that any matrix A induces a linear transformation $T_A(\underline{x}) = Ax^T$ and the matrix representation of T_A is simply A.
Now, to show P is the projection matrix, we need to establish the definition of the projection map. That is to show that given any vector $\underline{b} = (b_1, \ldots, b_n) \in \mathbb{R}^n$,

$$T_P(\underline{b}) = Pb^T \text{ lies in } W$$

and for any vector $\underline{x} \in W$, $b^T - Pb^T$ is orthogonal to $\underline{x}$. (view them as column vectors in $\mathbb{R}^n$)

So let $\underline{x} \in W$. Since $\text{col}(P) = W$, $\exists \underline{c} \in \mathbb{R}^n$ such that $Pc^T = x^T$.
Now,

$$\begin{aligned}
\langle b^T - Pb^T, Pc^T \rangle &= (b^T - Pb^T)^T Pc^T \\
&\qquad \text{(converting the dot product of real vectors to matrix multiplication)} \\
&= [(b^T)^T - (Pb^T)^T]Pc^T \\
&= (\underline{b} - \underline{b}P^T)Pc^T \\
&= b(I_{n\times n} - P^T)Pc^T \\
&= b(P - P^T P)c^T = b(P - P^2)c^T \qquad \text{(since } P \text{ is symmetric \& idempotent)} \\
&= b(P - P)c^T = [0]
\end{aligned}$$

Hence, $(b^T - Pb^T) \cdot (Pc^T) = 0$ when viewed as column vectors and hence orthogonal.

Hence $T_P = \text{proj}_W$ and since $T_P(\underline{x}) = Px^T$, the matrix representation (projection matrix) of T_P is P itself! ■

Notation: If we have two vectors $\underline{x} = (x_1, \ldots, x_n)$ and $\underline{y} = (y_1, \ldots, y_n)$ in $\mathbb{R}^n$, then

$$\langle \underline{x}, \underline{y} \rangle = \sum_{i=1}^{n} x_i y_i.$$

We can translate (or view) it in the language of matrix multiplication. So

$$\langle \underline{x}, \underline{y} \rangle = xy^T = \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} x_i y_i \end{bmatrix}_{1\times 1}.$$

Now, had we viewed $\underline{x}$ and $\underline{y}$ as column vectors $\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ and $\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$ in $\mathbb{R}^n$, then

$$\langle \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \rangle = x^T y = \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

would have been the translation or equivalent format with respect to the notion of matrix multiplication.

Example 3.3.1. Find the projection of an arbitrary vector $\underline{b} \in \mathbb{R}^3$ on the subspace W spanned by $\left\{ \begin{bmatrix} \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{bmatrix}, \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{bmatrix} \right\}$.

Answer: Let $\underline{b} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$ and let A $= \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & 0 \end{bmatrix}$, then $A^T A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I_2$.

The projection matrix $P = A(A^T A)^{-1} A^T = A(I)^{-1} A^T = AA^T = \begin{bmatrix} \frac{5}{6} & \frac{1}{6} & \frac{1}{3} \\ \frac{1}{6} & \frac{5}{6} & -\frac{1}{3} \\ \frac{1}{3} & -\frac{1}{3} & \frac{1}{3} \end{bmatrix}$.

Therefore,

$$\text{proj}_W \, \underline{b} = P\underline{b} = \begin{bmatrix} \frac{5}{6} & \frac{1}{6} & \frac{1}{3} \\ \frac{1}{6} & \frac{5}{6} & -\frac{1}{3} \\ \frac{1}{3} & -\frac{1}{3} & \frac{1}{3} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \frac{1}{6} \begin{bmatrix} 5b_1 + b_2 + 2b_3 \\ b_1 + 5b_2 - 2b_3 \\ 2b_1 - 2b_2 + 2b_3 \end{bmatrix} = b_W.$$

Example 3.3.2. True or False: Two different matrices may be projection matrices for the same subspace $W \subseteq \mathbb{R}^n$.

False. Let P and P' be two different projection matrices for $W \subseteq \mathbb{R}^n$, a subspace. Let $\underline{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$. We know $\text{proj}_W \underline{x} \in W$ and it's unique. That means for all $\underline{x} \in \mathbb{R}^n$, $Px^T = P'x^T \implies (P - P')x^T = 0 \implies$ $P - P'$ is the zero matrix, which means $P = P'$.

Definition 3.3.2. A real square matrix $A_{n\times n}$ is called orthogonal if $A^TA = AA^T = I$.

Remark 3.3.2. Soon we will see an analogous version of orthogonal matrices with entries from $\mathbb{C}$.

Remark 3.3.3. If A is orthogonal, then its columns form an orthonormal basis of $\mathbb{R}^n$ and so does its rows. Note that A is orthogonal if and only if A^T is orthogonal.

Definition 3.3.3. A linear map $T \in \mathcal{L}(V)$ on an inner product space is called an orthogonal linear map if $\langle T(v), T(w)\rangle = \langle v, w\rangle \quad \forall v, w \in V$.

Remark 3.3.4. In general an orthogonal linear map T is a linear map from any inner product space V to an inner product space W such that for all $v, w \in V$ we have $\langle T(v), T(w)\rangle_W = \langle v, w\rangle_V$

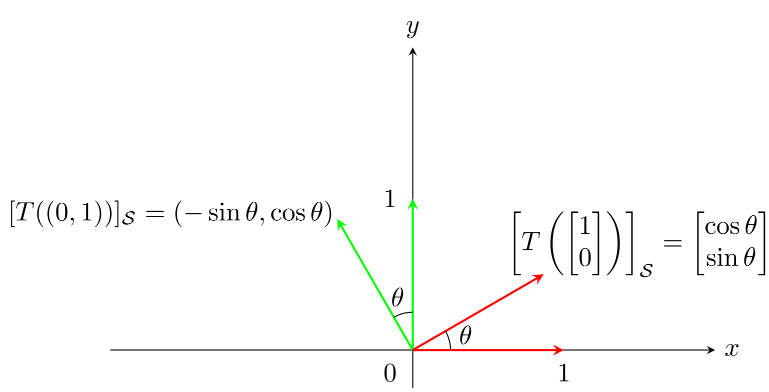
Remark 3.3.5. Clearly any orthogonal linear map T preserves norm. That is for all $v \in V$ we have that $\|v\| = \|T(v)\|$

Proposition 3.3.1. *$T \in \mathcal{L}(\mathbb{R}^n)$ is an orthogonal linear map* iff *its matrix representation is orthogonal.*

Proof. Assignment #4 ∎

Example 3.3.3. Consider the Euclidean inner product space $\mathbb{R}^2$. Then the linear map rotation (counter-clockwise) by angle $0 < \theta < \pi$ is an orthogonal linear map.

Answer: Let $\mathcal{S} = \langle (1,0), (0,1)\rangle$ be an ordered basis. We can compute the matrix representation $R_\mathcal{S}$ relative to ordered basis $\mathcal{S}$ as follows:



Note that for instance $(\cos(\theta), \sin(\theta)) = \cos(\theta)(1,0) + \sin(\theta)(0,1)$ so indeed $\big[T((1,0))\big]_\mathcal{S} = (\cos(\theta), \sin(\theta))$. Then $R_\mathcal{S} = \begin{bmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{bmatrix}$. It is easy to compute that

$$R_\mathcal{S}R_\mathcal{S}^T = \begin{bmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{bmatrix}\begin{bmatrix}\cos(\theta) & \sin(\theta)\\ -\sin(\theta) & \cos(\theta)\end{bmatrix} = \begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix} = \begin{bmatrix}\cos(\theta) & \sin(\theta)\\ -\sin(\theta) & \cos(\theta)\end{bmatrix}\begin{bmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{bmatrix} = R_\mathcal{S}^TR_\mathcal{S}$$

So $R_{\mathcal{S}}$ is orthogonal. By Proposition 3.3.1, the rotation map is an orthogonal linear map.

4 Eigenvalue/Eigenvector

4.1 Introduction

Motivation: There were two main applications that led to the discovery of such notions. The first was Markov Chains (Population Distribution), which will be studied in the last assignment, and the second was the Fibonacci sequence.

Recall Fibonacci sequence $(F_n)_{n\in\omega}$ is a recursively defined sequence with

$$F_0 = 0 \quad \text{and} \quad F_1 = 1 \quad \text{and} \quad \forall\, k \geq 2 \quad F_k = F_{k-1} + F_{k-2}.$$

Hence, if true, we can compute such terms using the notion of matrices.

Let $A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$ and let the vector $\underline{x}_{k-1} = \begin{bmatrix} F_{k-1} \\ F_{k-2} \end{bmatrix}$ represent the $k-1^{th}$ and $k-2^{th}$ term of the sequence. Then

$$A\underline{x}_{k-1} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} F_{k-1} \\ F_{k-2} \end{bmatrix} = \begin{bmatrix} F_{k-1} + F_{k-2} \\ F_{k-1} \end{bmatrix} = \begin{bmatrix} F_k \\ F_{k-1} \end{bmatrix} \quad \text{for } k \geq 2.$$

So $A\underline{x}_1 = \underline{x}_2 \implies x_3 = A\underline{x}_2 = A(A\underline{x}_1) = A^2\underline{x}_1$. Hence, $\underline{x}_n = A^{n-1}\underline{x}_1$.

So computing A^n gives rise to the $(n+1)^{th}$ term of the sequence. Luckily, A is of size 2×2 and its entries are simply 0 and 1. But the computation would have been cumbersome had it been the case that A was comprised of more complicated entries and of bigger size.

Definition 4.1.1 (Eigenvalue/Eigenvector)**.** Let $A \in \mathcal{M}_{n\times n}(\mathbb{F})$ where $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$. $\lambda \in \mathbb{F}$ is called an **eigenvalue** of A if there exists a nonzero vector $v = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} \in \mathbb{F}^n$ such that $Av = \lambda v$. Such nonzero vector v is called an **eigenvector** of A (corresponding to the eigenvalue λ).

Proposition 4.1.1 (properties of eigenvectors)**.**

I) For any $k \in \mathbb{N}$, $A^k v = \lambda^k v$.

II) If A is an invertible matrix and λ is an eigenvalue of A, then $\lambda \neq 0$ and $\frac{1}{\lambda}$ is an eigenvalue of A^{-1} corresponding to the same eigenvector of λ.

*III) Let $E_\lambda = \{\mathbf{0}\} \cup \{$eigenvectors associated with $\lambda\}$, then E_λ is a subspace of the n-space! ($\mathbb{C}^n$ or $\mathbb{R}^n$) E_λ is called the **eigenspace** of λ.*

Proof. II) Let A be an invertible matrix. If λ is 0 as an eigenvalue of A (by properties of vector space) then there is some nonzero $v \in \mathbb{F}^n$ such that $Av = \lambda v = 0 \cdot v = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}_{n \times 1}$ is the zero vector.

But A being invertible $\implies$ the only solution to the homogeneous system $A\underline{x} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$ is the trivial solution (i.e. $\underline{x} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$). But this contradicts the definition of eigenvector. So 0 must not be an eigenvalue of A.

Now $Av = \lambda v \implies A^{-1}Av = A^{-1}(\lambda v) \iff v = v = \lambda A^{-1} v \iff \frac{1}{\lambda} v = A^{-1} v$.
Note $\frac{1}{\lambda}$ represents the multiplicative inverse of $\lambda \in \mathbb{F}$! ∎

Exercise 4.1.1. Prove property I) and III)

Now we can make an observation that if we had a basis of $\mathbb{F}^n$ consisting of just eigenvectors we could facilitate the computation of $A^k v$! Recall $A^k = \underbrace{A \dots A}_{k\text{-times}}$ simply represents the transformation that is composed with itself k-many times.

Now let us see how we can simplify the computation of $A^k v$ in this case.

Let $\{b_1, b_2, \dots, b_n\}$ be a basis of my n-space consisting of just eigenvectors, each b_i corresponding to the eigenvalue λ_i respectively. Now given $v \in \mathbb{F}^n$, we have $v = \sum_{i=1}^{n} \alpha_i b_i$ where $\alpha_i \in \mathbb{F}$. Then

$$Av = A\left(\sum_{i=1}^{n} \alpha_i b_i\right) = \sum_{i=1}^{n} \alpha_i A b_i = \sum_{i=1}^{n} \alpha_i \lambda_i b_i$$

Note here b_i and v_i are viewed as $n \times 1$ matrices!

Hence by property I of the Proposition 4.1.1

$$A^k v = \alpha_1 \lambda_1^k b_1 + \alpha_2 \lambda_2^k b_2 + \dots + \alpha_n \lambda_n^k b_n$$

But now we should ask ourselves how does this linear map behave as $k \to \infty$? Note if $\exists\ j$ such that the norm of λ_j, $\|\lambda_j\| = |\lambda_j| > 1$ (the usual Euclidean norm), then $\lim_{k \to \infty} \|A^k v\| = \infty$. In such a case, we call A an **unstable linear map**. If $|\lambda_i| < 1\ \forall\ 1 \le i \le n$, then clearly $\lim_{k \to \infty} \|A^k v\| = 0$. A is said to be a **stable linear map**.

If $\max\{|\lambda_i| : 1 \le i \le n\} = 1$, without loss of generality assume $|\lambda_1| = 1 > |\lambda_i|$ for all $i > 1$, then A is said to be **neutrally stable** for $A^k v$ is approximately $\alpha_1 \lambda_1^k b_1$. That is because

$$A^k v = \alpha_1\lambda_1^k b_1 + \alpha_2\lambda_2^k b_2 + \cdots + \alpha_n\lambda_n^k b_n = \lambda_1^k[\alpha_1 b_1 + \underbrace{\alpha_2\Big(\frac{\lambda_2}{\lambda_1}\Big)^k b_2 + \cdots + \alpha_n\Big(\frac{\lambda_n}{\lambda_1}\Big)^k b_n}_{\text{The NORM of this part goes to 0 as } k\to\infty}]$$

$\therefore \left\|A^k v - \alpha_1\lambda_1^k b_1\right\|$ is as small as possible compared with $\left\|A^k v\right\|$

Soon we will learn a method called "Power Method" that will allow us to compute the eigenvalue of maximum norm. Of course, sometimes there are very easy approaches to compute the λ_i of maximum norm if one is quite dominant. We can compare the components of $A^{k+1}v$ to those of $A^k v$ for quite small values of k.

Note: I can make quite challenging/tricky types of questions based on this simple notion which only would require basic algebra. Hence it's extremely important to understand the overall picture!

4.2 Computing Eigenvector

Recall to compute eigenvector v we need to find a scalar $\lambda \in \mathbb{F}$ such that

$$Av = \lambda v \iff Av - \lambda v = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}_{n\times 1} \iff (A - \lambda I)v = \hat{0} \quad (*)$$

Note that if the matrix $A - \lambda I$ is invertible (i.e. $\det(A-\lambda I) \neq 0$) then the only solution $v = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}$ to the homogeneous system $(*)$ would be the trivial solution $\begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$, but such a solution cannot be considered as an eigenvector.

Definition 4.2.1 (Characteristic Polynomial)**.** If A is an $n \times n$ matrix, the determinant $\det(A - \lambda I)$ is a polynomial of degree n of variable λ. This polynomial is called the **characteristic polynomial** of A.

Hence, we want $\det(A - \lambda I) = 0$. So the roots of the characteristic polynomial gives us the eigenvalues, say λ_i. Finding the nullspace of $A - \lambda_i I$ gives us the set of eigenvectors associated with the eigenvalue λ_i.

Remark 4.2.1. Note that the eigenspace E_{λ_i} is just the nullspace of $A - \lambda_i I$. We can find a basis for it by solving $(A - \lambda_i I)v = 0$.

Remark 4.2.2. The main difference between complex and real eigenvalues is that not every matrix or equivalently a linear transformation yields a real eigenvalue. The following example shows that there are real matrices with strictly complex eigenvalues.

Example 4.2.1. Consider the matrix $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ which literally represents the rotation(counterclockwise) linear map by angle $\frac{\pi}{2}$ by example 3.3.3. Then $|A - \lambda I| = \left| \begin{bmatrix} -\lambda & -1 \\ 1 & -\lambda \end{bmatrix} \right| = \lambda^2 + 1$, which has no real solutions. However, we do have complex eigenvalues; in particular, $\lambda_1 = i$ and $\lambda_2 = -i$. So $\lambda^2 + 1 = (\lambda - i)(\lambda + i)$.

Remark 4.2.3. The fundamental theorem of algebra states every polynomial of degree n has exactly n roots counting multiplicity of each root (zero). Hence we always get complex-valued eigenvalues.

Now note as usual, since there is an one-to-one correspondence between matrices and linear transformations between finite-dimensional spaces, we can define the same notion of eigenvalue/eigenvectors for any linear transformations.

Definition 4.2.2. Let $T: V \to V$. $\lambda \in \mathbb{F}$ is an **eigenvalue** of T, then there exists a nonzero vector v such that $Tv = \lambda v$. This v is said to be an **eigenvector** (corresponding to λ).

Remark 4.2.4. First, we are going to study all the theorems and notions associated with matrices with respect to both $\mathcal{M}_{n\times n}(\mathbb{R})$ and $\mathcal{M}_{n\times n}(\mathbb{C})$. Then we are going to study the analogous notions with respect to linear transformations.

4.3 Diagonalization

Motivation: In order to compute A^k, one often way is to diagonalize the matrix A.

Definition 4.3.1 (Diagonalizable)**.** A square matrix A is said to be diagonalizable if there exists an invertible matrix $C_{n\times n}$ and a diagonal matrix D such that $C^{-1}AC = D$.

Let us revisit some theorems that are covered in MATA22 without proving them. Note that the following theorem also works if $A \in M_{n\times n}(\mathbb{C})$.

Theorem 4.3.1. *Let A be an $n \times n$ matrix. Let $\lambda_1, \dots, \lambda_n$ be some scalars (some possibly being complex numbers), and $v_1, \dots, v_n$ be non-zero vectors. Let C be an $n \times n$ matrix whose j-th column is the column vector v_j and*

$$D = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix}$$

then $AC = CD$ if and only if $\lambda_1, \dots, \lambda_n$ are eigenvalues of A with corresponding eigenvectors v_i respectively.

Proof. We show the conclusion simultaneously:

$$\begin{aligned} CD &= \begin{bmatrix} | & & | \\ v_1 & \dots & v_n \\ | & & | \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix} \\ &= \begin{bmatrix} | & & | \\ \lambda_1 v_1 & \dots & \lambda_n v_n \\ | & & | \end{bmatrix} \\ &= \begin{bmatrix} | & & | \\ Av_1 & \dots & Av_n \\ | & & | \end{bmatrix} \\ &= A \begin{bmatrix} | & & | \\ v_1 & \dots & v_n \\ | & & | \end{bmatrix} \\ &= AC \end{aligned}$$

if and only if $Av_i = \lambda_i v_i$ for all $1 \le i \le n$ ∎

Remark 4.3.1. Let $A \in \mathcal{M}_{n \times n}(\mathbb{C})$ be diagonalizable. Then we obtain (by the definition of diagonalizability) $AC = CD$. This implies that $A = CDC^{-1}$. Hence for any $k \in \mathbb{N}$,

$$\begin{aligned} A^k &= (CDC^{-1})^k \\ &= \underbrace{(CDC^{-1})(CDC^{-1}) \dots (CDC^{-1})}_{k\text{-times}} \\ \implies A^k &= CD^kC^{-1} \end{aligned}$$

Remark 4.3.2. The remark above should convince us of our original claim. Thereby we can easily compute A^k using a diagonal matrix D

Remark 4.3.3. Recall a matrix $C_{n \times n}$ is invertible if and only if $\text{rank}(C) = n$ if and only if column vectors of C from a basis of a n-space.

Now we proceed to establish a sufficient but NOT necessary condition for diagonalization.

Corollary 4.3.1.1. *Let A be an $n \times n$ matrix with $r \leq n$ distinct eigenvalues $\lambda_1, \ldots, \lambda_r$, (some may be complex values) then the corresponding eigenvectors are linearly independent. In particular, if $r = n$, then A is diagonalizable.*

Proof. We proceed to prove this by contradiction. Assume on the contrary that $\{v_1, \ldots, v_n\}$ is linearly dependent. WLOG, let n be the first index such that v_n is dependent in the set $\{v_1, \ldots, v_{n-1}, v_n\}$ and $\{v_1, \ldots, v_{n-1}\}$ is linearly independent. Hence there are some non-zero scalars $\alpha_i \in \mathbb{C}$ such that

$$v_n = \alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_{n-1} v_{n-1} \qquad (\star)$$

We now manipulate $(\star)$ in two different ways using the notion of eigenvalues and eigenvectors

$$\lambda_n v_n = \lambda_n \alpha_1 v_1 + \cdots + \lambda_n \alpha_{n-1} v_{n-1} \qquad (1)$$

$$\begin{aligned} Av_n &= A\alpha_1 v_1 + A\alpha_2 v_2 + \cdots + A\alpha_{n-1} v_{n-1} \\ &= \alpha_1 A v_1 + \cdots + \alpha_{n-1} A v_{n-1} \\ &= \alpha_1 \lambda_1 v_1 + \cdots + \alpha_{n-1} \lambda_{n-1} v_{n-1} \qquad (2) \end{aligned}$$

We see that since the left sides of the equations (1) and (2) are equal, we must have that the right hand sides are equal as well. Thus taking (1) $-$ (2), we see that

$$\begin{aligned} 0 &= (\lambda_n \alpha_1 v_1 - \alpha_1 \lambda_1 v_1) + (\lambda_n \alpha_2 v_2 - \lambda_2 \alpha_2 v_2) + \cdots + (\lambda_n \alpha_{n-1} v_{n-1} - \alpha_{n-1} \lambda_{n-1} v_{n-1}) \\ 0 &= (\lambda_n - \lambda_1)\alpha_1 v_1 + (\lambda_n - \lambda_2)\alpha_2 v_2 + \cdots + (\lambda_n - \lambda_{n-1})\alpha_{n-1} v_{n-1} \end{aligned}$$

However since all λ_i are distinct, we have that $\lambda_n - \lambda_i \neq 0$ for all $1 \leq i \leq n-1$. Since at least one of α_i is also non-zero, say α_1, then $\alpha_1(\lambda_n - \lambda_1) \neq 0$ which in turn contradicts the fact that the set $\{v_1, \ldots, v_{n-1}\}$ is linearly independent. Therefore we must have that $\{v_1, \ldots, v_n\}$ is linearly independent and by our previous remarks and theorem, A is diagonalizable. ■

Theorem 4.3.2 (Equivalent conditions to diagonalizability)**.**
The following are equivalent:

1. *A is diagonalizable.*
2. *The space contains a basis consisting of just eigenvectors of A.*
3. *For each eigenvalue λ of A, the geometric multiplicity is the same as the algebraic multiplicity.*

Remark 4.3.4. The **algebraic multiplicity** is the multiplicity of an eigenvalue λ as a root of the characteristic polynomial $p(\lambda) = \det(A - \lambda I)$. The **geometric multiplicity** of an eigenvalue λ is the simply the dimension of the corresponding eigenspace E_λ.

Remark 4.3.5. For any eigenvalue λ **geometric multiplicity** $\leq$ **algebraic multiplicity** of λ

The following Theorem is extremely important for characterizing diagonalizable matrices.

Theorem 4.3.3 (Spectral Theorem, real version)**.** *Let $A \in \mathcal{M}_{n\times n}(\mathbb{R})$. If A is symmetric, then A is diagonalizable by an orthogonal matrix.*

Remark 4.3.6. Soon we will see the spectral theorem for the matrices with complex entries.

Remark 4.3.7. Let A be a symmetric matrix with eigenvalues $\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$, not necessarily distinct. Then the **spectral decomposition** of A is

$$A = b_1 b_1^T \lambda_1 + b_2 b_2^T \lambda_2 + \cdots + b_n b_n^T \lambda_n \tag{1}$$

where $C = \begin{bmatrix} | & | & \cdots & | \\ b_1 & b_2 & \cdots & b_n \\ | & | & \cdots & | \end{bmatrix}$ is the orthogonal matrix that diagonalizes the matrix A.

Exercise 4.3.1. Find and verify the spectral decomposition of A where $A = \begin{bmatrix} 3 & -2 \\ -2 & 0 \end{bmatrix}$.

Now let us study the notions that will allow us to establish the above spectral theorem in general.

Definition 4.3.2 (Conjugate Transpose - Adjoint)**.** Let $A \in \mathcal{M}_{n\times n}(\mathbb{C})$ where $A = [a_{ij}]$. Then $A^* = [\overline{a_{ij}}]^T = [\overline{a_{ji}}]$.

Example 4.3.1. Consider $A = \begin{bmatrix} 1 & 0 & i \\ 2 & i-1 & 1 \\ 0 & 2 & \sqrt{2}i \end{bmatrix}$, then $A^* = \begin{bmatrix} 1 & 2 & 0 \\ 0 & -i-1 & 2 \\ -i & 1 & -\sqrt{2}i \end{bmatrix}$.

Definition 4.3.3 (Hermitian)**.** $A \in \mathcal{M}_{n\times n}(\mathbb{C})$ is called a Hermitian matrix if $A = A^*$.

Definition 4.3.4 (Unitary)**.** $A \in \mathcal{M}_{n\times n}(\mathbb{C})$ is called a Unitary matrix if $AA^* = A^*A = I$.

We have the following properties of conjugate transpose:

Proposition 4.3.1. *Given $A \in M_{n\times n}(\mathbb{C})$,*

I) $(A^*)^* = A$

II) $(A+B)^* = A^* + B^*$

III) $\forall z \in \mathbb{C}\ (zA)^* = \overline{z}A^*$

IV) $\forall B \in M_{n\times n}(\mathbb{C}),\ (AB)^* = B^*A^*$

Proof. III)

$$\begin{aligned} zA = [za_{ij}] \implies (zA)^* &= [\overline{za_{ij}}]^T \\ &= [\overline{z}\ \overline{a_{ij}}]^T && \text{(properties of transpose)} \\ &= \overline{z}[\overline{a_{ij}}]^T \\ &= \overline{z}A^* \end{aligned}$$

■

Exercise 4.3.2. Prove *I), II)* and *IV)* of Proposition 4.3.1

Remark 4.3.8. Hermitian and unitary matrices are the "complex version" of symmetric and orthogonal matrices respectively. You can prove this easily by definition.

Now we can restate Theorem 4.3.3 in general:

Theorem 4.3.4 (Spectral Theorem)**.** *Every Hermitian matrix A is unitarily diagonalizable.*

Remark 4.3.9. Note that the transpose of eigenvectors b_i in the Spectral Decomposition (1) of Theorem 4.3.3 will be replaced by $\overline{b_i}^T$

Example 4.3.2. Let $A = \begin{bmatrix} 1 & 0 & i \\ 0 & 2 & 0 \\ -i & 0 & 1 \end{bmatrix}$. Is A diagonalizable? If so, find an invertible matrix that diagonalizes A.

Answer: Note $A = A^* = \begin{bmatrix} 1 & 0 & i \\ 0 & 2 & 0 \\ -i & 0 & 1 \end{bmatrix}$. So by the spectral theorem, it's diagonalizable by a unitary matrix.

Let us first find the eigenvalues.

The characteristic polynomial $p(\lambda)$ is given by:

$$p(\lambda) = \det(A - \lambda I) = \begin{vmatrix} 1-\lambda & 0 & i \\ 0 & 2-\lambda & 0 \\ -i & 0 & 1-\lambda \end{vmatrix} = (2-\lambda)\begin{vmatrix} 1-\lambda & i \\ -i & 1-\lambda \end{vmatrix} = -\lambda(2-\lambda)^2$$

$\therefore$ The eigenvalues are $\lambda_1 = 0$ and $\lambda_2 = 2$.

First note in general algebraic multiplicity $\geq$ geometric multiplicity of an eigenvalue.
If the algebraic multiplicity of an eigenvalue is 1, its geometric multiplicity of that eigenvalue λ, regardless of whether A is diagonalizable or not, has to be 1. For if the geometric multiplicity is 0 $\Rightarrow E_\lambda = \{0\}$ and we know that an eigenvector is a nonzero vector.

Now we solve for the eigenvector associated with λ_1:

$$A - \lambda_1 I = \begin{bmatrix} 1 & 0 & i \\ 0 & 2 & 0 \\ -i & 0 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & i \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$\therefore E_{\lambda_1} = \text{span}\left\{ \begin{bmatrix} -i \\ 0 \\ 1 \end{bmatrix} \right\}.$

Now to find the eigenvectors corresponding to the eigenvalue $\lambda_2 = 2$, which has algebraic multiplicity 2. Note since A is Hermitian, it is diagonalizable. Hence the geometric multiplicity of λ_2 must also be 2.

$$A - \lambda_2 I = \begin{bmatrix} -1 & 0 & i \\ 0 & 0 & 0 \\ -i & 0 & -1 \end{bmatrix} \sim \begin{bmatrix} -1 & 0 & i \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$E_{\lambda_2} = \text{span}\left\{ \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} i \\ 0 \\ 1 \end{bmatrix} \right\}$$

Now note that the matrix $C = \begin{bmatrix} -i & 0 & i \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$ (where the columns are just basis vectors of eigenspaces) does diagonalize the matrix A! Also, the columns are orthogonal!

Note $\left\langle \begin{bmatrix} -i \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} i \\ 0 \\ 1 \end{bmatrix} \right\rangle = \begin{bmatrix} i & 0 & 1 \end{bmatrix} \begin{bmatrix} i \\ 0 \\ 1 \end{bmatrix} = 0$, so simply as vectors $(\overline{-i}) \cdot i + 0 \cdot 0 + 1 \cdot 1 = 0$.

We can find the diagonalization by computation:

$$A = \begin{bmatrix} -i & 0 & i \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} -i & 0 & i \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}^{-1}$$

Note the order of columns of C corresponds to the order of eigenvalues in the diagonal of the diagonal matrix.

However, the matrix C is **NOT** a unitary matrix. Note that its first and third columns have norm $\|(1, 0, i)\| = \sqrt{2}$. Thus, the unitary matrix U which diagonalizes A is $\begin{bmatrix} -\frac{i}{\sqrt{2}} & 0 & \frac{i}{\sqrt{2}} \\ 0 & 1 & 0 \\ \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \end{bmatrix}$

Remark 4.3.10. A is diagonalizable means that there exists an invertible matrix C such that $C^{-1}AC$ is a diagonal matrix. Hence, there could be infinitely many C that diagonalize A (i.e. We can multiply C by any nonzero scalar). However, being Hermitian allows us to find a unitary matrix U.

Now let us prove the Spectral Theorem. To do so, we first introduce a lemma that will be invoked in proving the theorem.

Lemma 4.3.1 (Schur's Lemma)**.**
Let $A \in \mathcal{M}_{n\times n}(\mathbb{C})$. There exists a unitary matrix U such that $U^*AU = T$ where T is an upper triangular matrix.

Now we are equipped to prove the Spectral Theorem for matrices.

Theorem 4.3.4 (Spectral Theorem)**.** *Every Hermitian matrix A is unitarily diagonalizable.*

Proof. Let $A \in \mathcal{M}_{n\times n}(\mathbb{C})$ be a Hermitian matrix. By Schur's Lemma, $\exists$ a unitary matrix U such that $U^*AU = D$ where D is an upper triangular matrix with all the entries below the diagonal being 0.

Note that $D^* = (U^*AU)^* = U^*A^*(U^*)^* = U^*A^*U = U^*AU = D$ since A is Hermitian and we applied the properties of the conjugate transpose. Now the transpose of an upper triangular matrix is a lower triangular matrix, and $D^* = D \implies D$ is both upper triangular and lower triangular $\implies D$ is diagonal and its diagonal entries are eigenvalues of A which are actually all real numbers. ■

Remark 4.3.11. The diagonal entries of a Hermitian matrix are real.

Corollary 4.3.4.1. *The eigenvalues of a Hermitian matrix are all real.*

Exercise 4.3.3. True or False: Suppose $A \in \mathcal{M}_{n\times n}(\mathbb{C})$ is diagonalizable. Then A is Hermitian.

We saw earlier that the proof of the theorem is by Schur's Lemma. So let us prove the lemma. However, let us first establish another important theorem.

Remark 4.3.12. From now on, given $\underline{x}$ and $\underline{y} \in \mathbb{C}^n$, we should always view them as column vectors:

$$\underline{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \quad \underline{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

since the standard Hermitian inner product on $\mathbb{C}^n$ is linear in the second slot:

$$\langle \underline{x}, \underline{y} \rangle = \sum_{i=1}^{n} \overline{x_i} y_i = x^* y$$

Theorem 4.3.5. *Let $A \in \mathcal{M}_{n\times n}(\mathbb{C})$ be a Hermitian matrix. The eigenvectors of A corresponding to distinct eigenvalues are orthogonal.*

Proof. Let $\{\lambda_1, \ldots, \lambda_k\}$ be eigenvalues of A where $k \leq n$ with corresponding eigenvectors $\{v_1, \ldots, v_k\}= X$. Consider two arbitrary eigenvectors $v, w \in X$ with corresponding eigenvalues λ_1, λ_2 respectively. We need to show $\langle v, w\rangle = 0$.
Note $\lambda_2(v^*w) = v^*\lambda_2 w$ since λ_2 is a scalar and we can view v^* and w as row matrix and column matrix respectively. Now

$$\begin{aligned}
\lambda_2(v^*w) &= v^*(\lambda_2 w) = v^*Aw && \text{(since } Aw = \lambda_2 w)\\
&= (v^*A)w = (A^*v)^*w\\
&= (Av)^*w && \text{(since } A* = A)\\
&= (\lambda_1 v)^*w && \text{(since } Av = \lambda_1 v)\\
&= \overline{\lambda_1} v^*w && \text{(properties of conjugate)}\\
&= \lambda_1 v^*w && \text{(Corollary 4.3.4.1)}
\end{aligned}$$

But then

$$\lambda_2 v^*w = \lambda_1 v^*w \iff (\lambda_2 - \lambda_1)v^*w = 0$$

Since $\lambda_1 \neq \lambda_2 \implies \lambda_2 - \lambda_1 \neq 0$, and $v * w \in \mathbb{C}$, $\mathbb{C}$ is a field, we have

$$v^*w = 0 = \langle v, w\rangle$$

■

Now we will prove Schur's Lemma.

Lemma 4.3.1 (Schur's Lemma)**.**
Let $A \in \mathcal{M}_{n\times n}(\mathbb{C})$. There exists a unitary matrix U such that $U^*AU = T$ where T is an upper triangular matrix.

Proof. We prove this lemma by induction on the size of the matrix A.
Let $n = 1$. Say $A = [\alpha]$ where $\alpha \in \mathbb{C}$. Then clearly taking $U = U^* = [1]$ we have an upper triangular matrix U^*AU.
Now suppose that Schur's Lemma holds for all matrices of size at most $(n-1)\times(n-1)$.
Let λ_1 be an eigenvalue of A associated with an eigenvector v_1. (Why does there even exist an eigenvalue?)
The purpose/intuition behind this proof is to use the inductive hypothesis to actually construct a unitary matrix with the mentioned properties.
So, to somehow build a unitary matrix we need to have that the columns are orthonormal! Hence, why not extend the linearly independent set $\{v_1\}$ (since eigenvectors are nonzero) to a basis of the vector space $\mathbb{C}^n$ over $\mathbb{C}$.

Let $\{v_1, v_2, \ldots, v_n\}$ be such an extension. Now apply the Gram-Schmidt process, fixing v_1 as the first element and normalizing all the basis elements to obtain the following orthonormal basis.

$$\left\{ v_1 = \begin{bmatrix} v_{11} \\ v_{21} \\ \vdots \\ v_{n1} \end{bmatrix}, \cdots, v_n = \begin{bmatrix} v_{1n} \\ v_{2n} \\ \vdots \\ v_{nn} \end{bmatrix} \right\}$$

Now consider the unitary matrix $U_1 = \begin{bmatrix} | & & | \\ v_1 & \cdots & v_n \\ | & & | \end{bmatrix}$.

Intuitively this should not be the final unitary matrix we want since we have not yet utilized the inductive hypothesis.

Note $U_1^* = \begin{bmatrix} — & \overline{v_1} & — \\ & \vdots & \\ — & \overline{v_n} & — \end{bmatrix}$, *and its ith row is orthogonal to the jth column of U_1 for $i \neq j$.*

Now consider the matrix $U_1^* A U_1 = \begin{bmatrix} — & \overline{v_1} & — \\ & \vdots & \\ — & \overline{v_n} & — \end{bmatrix} A \begin{bmatrix} | & & | \\ v_1 & \cdots & v_n \\ | & & | \end{bmatrix} = \begin{bmatrix} — & \overline{v_1} & — \\ & \vdots & \\ — & \overline{v_n} & — \end{bmatrix} \begin{bmatrix} | & & | \\ Av_1 & \cdots & Av_n \\ | & & | \end{bmatrix}$.

Note since $Av_1 = A \begin{bmatrix} v_{11} \\ v_{21} \\ \vdots \\ v_{n1} \end{bmatrix} = \lambda_1 v_1 = \begin{bmatrix} \lambda_1 v_{11} \\ \vdots \\ \lambda_1 v_{n1} \end{bmatrix}$, *and* $\langle v_1, v_1 \rangle = {v_1^*}^T v_1 = \begin{bmatrix} \overline{v_{11}} & \overline{v_{21}} & \cdots & \overline{v_{n1}} \end{bmatrix} \begin{bmatrix} v_{11} \\ v_{21} \\ \vdots \\ v_{n1} \end{bmatrix} = 1$, *and* $\forall\ i > 1,\ \langle v_i, v_1 \rangle = 0 = v_i^* v_1$. *Thus, the first column of $U_1^* A U_1$ will be*

$$\begin{bmatrix} \lambda_1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}_{n \times 1}$$

Now let us represent $U_1^* A U_1$ symbolically as

$$C = \left[\begin{array}{c|ccc} \lambda_1 & * & \cdots & * \\ \hline 0 & & & \\ \vdots & & \mathbf{A_1} & \\ 0 & & & \end{array}\right] \qquad (*)$$

It does not matter what the entries c_{1j} for $j \geq 2$ (which we replaced with $*$) and c_{ik} for $i \geq 2, k \geq 2$ are. Deleting the 1^{st} column and 1^{st} row of C, we obtain A_1 which is of size $(n-1) \times (n-1)$.
By the inductive hypothesis, there exists a unitary matrix ρ such that

$$\rho^* A_1 \rho = B,$$

where B is an upper triangular matrix.

Let $U_2 = \left[\begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ \hline 0 & & & \\ \vdots & & \rho & \\ 0 & & & \end{array}\right]$ $(\triangle)$*. Note* U_2 *is unitary, and* $U_2^* = \left[\begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ \hline 0 & & & \\ \vdots & & \rho^* & \\ 0 & & & \end{array}\right]$*. (Verify this!)*
Let $U = U_1 U_2$*. Then*

$$\begin{aligned} UU^* &= (U_1U_2)(U_1U_2)^* \\ &= U_1U_2U_2^*U_1^* \\ &= U_1IU_1^* && (U_2 \text{ is unitary}) \\ &= U_1U_1^* = I && (U_1 \text{ is unitary}) \\ &= U_2^*U_1^*U_1U_2 \\ &= (U_1U_2)^*(U_1U_2) \\ &= U^*U \end{aligned}$$

Now,

$$\begin{aligned} U^*AU &= (U_2^*U_1^*)A(U_1U_2) \\ &= U_2^*(U_1^*AU_1)U_2 \\ &= U_2^* \left[\begin{array}{c|ccc} \lambda_1 & * & \cdots & * \\ \hline 0 & & & \\ \vdots & & \mathbf{A_1} & \\ 0 & & & \end{array}\right] U_2 && (\text{by } (*)) \\ &= \left[\begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ \hline 0 & & & \\ \vdots & & \rho^* & \\ 0 & & & \end{array}\right] \left[\begin{array}{c|ccc} \lambda_1 & * & \cdots & * \\ \hline 0 & & & \\ \vdots & & \mathbf{A_1} & \\ 0 & & & \end{array}\right] \left[\begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ \hline 0 & & & \\ \vdots & & \rho & \\ 0 & & & \end{array}\right] && (\text{by } (\triangle)) \\ &= \left[\begin{array}{c|ccc} \lambda_1 & * & \cdots & * \\ \hline 0 & & & \\ \vdots & & \mathbf{B} & \\ 0 & & & \end{array}\right] && (\text{by computation}) \end{aligned}$$

which is upper triangular, since B *is upper triangular.* ∎

Exercise 4.3.4. You should convince yourself in the above proof that:

U_2 is unitary, and $U_2^*CU_2 = \left[\begin{array}{c|ccc} \lambda_1 & * & \cdots & * \\ \hline 0 & & & \\ \vdots & & \mathbf{B} & \\ 0 & & & \end{array}\right]$, even though they are true by direct computation.

Lastly let us characterize all unitarily diagonalizable matrices.

Definition 4.3.5 (Normal)**.** Let $A \in \mathcal{M}_{n\times n}(\mathbb{C})$. A is said to be **normal** if $A^*A = AA^*$.

Theorem 4.3.6. *$A \in \mathcal{M}_{n\times n}(\mathbb{C})$ is **unitarily diagonalizable** iff A is **normal**.*

Remark 4.3.13. Note that being unitarily diagonalizable is stronger than being diagonalizable and converse may not be true! Unitarily diagonalizable indicates the diagonalization occurs through a matrix whose columns are orthonormal, or simply put orthonormal basis. Whereas the latter case only depicts that the diagonalization is attained by just an **invertible** matrix. Even though the columns of that invertible matrix are comprised of **eigenvectors** of the matrix, however, they may not be orthogonal. The following theorem should convince us as to why this may be the case.

Theorem 4.3.7. *Let $A \in \mathcal{M}_{n\times n}(\mathbb{C})$ be a Hermitian matrix. Then the eigenvectors of A corresponding to distinct eigenvalues are orthogonal.*

Proof. Let v, w be two eigenvectors of A with corresponding eigenvalues λ_1 and λ_2 respectively. Then

$$(1)\langle \lambda_1 v, w\rangle = (\lambda_1 v)^* w = \overline{\lambda}_1 v^* w = \lambda_1 v^* w \qquad \text{(since eigenvalues of Hermitian matrix are real)}$$

$$(2)\langle \lambda_1 v, w\rangle = (\lambda_1 v)^* w = (Av)^* w = v^* A^* w = v^* A w = v^* \lambda_2 w = \lambda_2 v^* w \qquad \text{(since } A \text{ is Hermitian)}$$

Subtracting (1) by (2) we obtain that $(\lambda_1 - \lambda_2)(v^* w) = 0$. Note that $\lambda_1 - \lambda_2$ is a non-zero scalar in $\mathbb{C}$ since eigenvalues are distinct. Since $\mathbb{C}$ is a field that means the scalar $v^* w$ must be zero! Hence, we have that $\langle v, w\rangle = v^* w = 0$ ■

Exercise 4.3.5. Find a matrix A which is diagonalizable but not unitarily diagonalizable.

Example 4.3.3. Find all the values of $a \in \mathbb{C}$ such that $A = \begin{bmatrix} i & a \\ 2 & i \end{bmatrix}$ is unitarily diagonalizable.

Answer: By Theorem 4.3.6 we need to have $\begin{bmatrix} i & a \\ 2 & i \end{bmatrix} \begin{bmatrix} -i & 2 \\ \overline{a} & -i \end{bmatrix} = \begin{bmatrix} -i & 2 \\ \overline{a} & -i \end{bmatrix} \begin{bmatrix} i & a \\ 2 & i \end{bmatrix}$

Hence, we need to have that $a\overline{a} + 1 = 5$, as the entry in the first row first column of the product matrix, which reduces down to the equation $\|a\|^2 = |a|^2 = 4$. Therefore, $\{x + iy = z \in \mathbb{C} : z\overline{z} = x^2 + y^2 = 4\}$ is the solution set.

5 Operators on Inner Product Spaces

5.1 Self-Adjoint and Normal Operators

In this section, we are going to develop all the notions analogous to what we developed in the previous section which evolves from Definition 4.2.2

Definition 5.1.1 (Linear Functional)**.** Let V be a complex inner product space. Then any linear map $\psi : V \to \mathbb{C}$ is called a **linear functional.**

Remark 5.1.1. Note that we view the co-domain of our map ψ as a vector space $\mathbb{C}$ over itself!

Remark 5.1.2. As previously mentioned, by fixing any $v \in V$ we can induce a "natural" linear functional as follows: For all $w \in V$, $\psi_v(w) = \langle v, w\rangle$. We used the subscript to indicate the definition of ψ really depends on the vector v. So, for every $v \in V$ we can easily construct a linear functional. Note that if we are given an explicitly defined inner vector space, we can specifically pick a v in the space and consider its induced linear functional depending on what we want to show. I may/or may not test you on this idea. So solving such a question would require you to apply and utilize such manipulation and idea.

Now you might ponder, is Remark 5.1.2 capturing all the possible linear functionals that may exist on V or there are other linear functionals that cannot be really reflected by an inner product? The question is **negatively** answered by the following theorem for any **finite** dimensional inner product space.

Theorem 5.1.1 (Riesz Representation)**.** *Let V be a finite dimensional complex inner product space. Let ψ be a linear functional on V. Then there exists a* ***unique*** *$u \in V$ such that for all $v \in V$ we have that:*

$$\psi(v) = \langle u, v\rangle$$

Proof. Let $\{q_1, \cdots, q_n\}$ be an orthonormal basis of V. Take any $v \in V$. Then $v = \sum_{i=1}^{n} \alpha_i q_i = \sum_{i=1}^{n} \langle q_i, v\rangle q_i$. Now by linearity of ψ we have that $\psi(v) = \psi\big(\sum_{i=1}^{n} \langle q_i, v\rangle q_i\big) = \sum_{i=1}^{n} \psi(\langle q_i, v\rangle q_i) = \sum_{i=1}^{n} \langle q_i, v\rangle \psi(q_i)$. Now applying the linearity of the Hermitian inner product in the second slot we obtain that $\sum_{i=1}^{n} \langle \overline{\psi(q_i)} q_i, v\rangle = \langle \sum_{i=1}^{n} \overline{\psi(q_i)} q_i, v\rangle$. ■

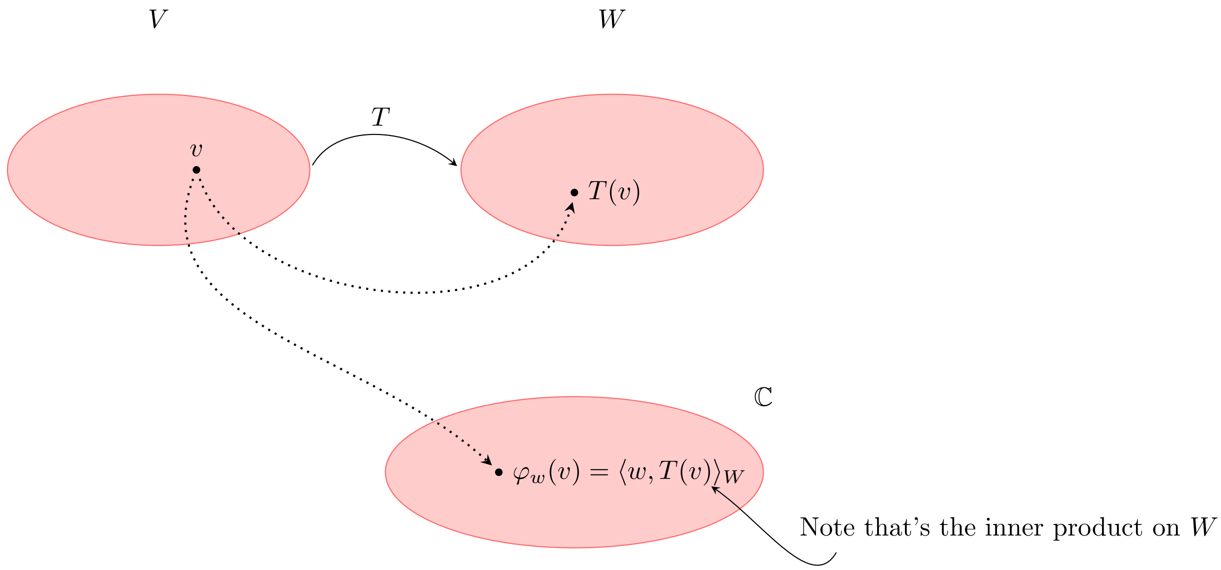
Now we are ready to develop and study the analogous notions that we saw for matrices for linear operators. We begin by introducing some definitions and properties.

Definition 5.1.2 (Adjoint)**.** Let V, W be complex inner product spaces and $T \in \mathcal{L}(V, W)$. The adjoint of T, denoted by T^*, is a function $T^* : W \to V$ such that for all $v \in V$ and $w \in W$,

$$\langle w, T(v)\rangle = \langle T^*(w), v\rangle$$

Remark 5.1.3. You should ask yourself, does such a map exist to begin with? The Riesz Representation Theorem ensures the existence of such a function.

Let $T : V \to W$ and fix $w \in W$.



So we can define a linear functional on V using the fixed vectors $w \in W$ and the linear map T as illustrated in the diagram above.

Hence we have two parameters. For every such pair, we could define a different linear functional. Since T is given and we want to show the adjoint of T exists, the only parameter of interest will be the arbitrary vectors $w \in W$. So $\forall v \in V,\ \varphi_w(v) = \langle w, T(v) \rangle$.

By Riesz Representation Theorem:

$$\text{There exists a unique } u_w \in V \text{ such that } \langle u_w, v \rangle = \varphi_w(v) = \langle w, T(v) \rangle \text{ for all } v \in V.$$

We use the subscript w since that unique vector u was dependent on the parameter w which defined the linear functional ϕ.

Hence, for every vector $w \in W$ we can define a different linear functional on V. Thereby we can define a map

$$\begin{aligned} T^* : W &\to V \\ w &\longmapsto u_w \end{aligned}$$

where u_w is the unique vector determined by the Riesz Representation Theorem, and we have

$$\langle w, T(v)\rangle = \langle T^*(w), v\rangle$$

which is what we need.

Remark 5.1.4. Given $T \in \mathcal{L}(V, W)$, the adjoint of map T^* is a linear map from W to V.

Exercise 5.1.1. Show $T^* \in \mathcal{L}(W, V)$.

As for matrices, the adjoint operation, which is reflecting what the conjugate transpose operator on matrices does. The adjoint T^* of a linear transformation T satisfies similar properties:

Proposition 5.1.1. *Let $T \in \mathcal{L}(V, W)$.*

I) $\forall S \in \mathcal{L}(V, W)$, $(S + T)^* = S^* + T^*$.

II) $(T^*)^* = T$.

III) $\forall \alpha \in \mathbb{C}$, $(\alpha T)^* = \overline{\alpha} T^*$.

IV) $\forall K \in \mathcal{L}(W, V)$, $(K \circ T)^* = T^* \circ K^*$.

Proof. I) Fix $v \in V$ and let $w \in W$ be arbitrary. First note $(S+T) \in \mathcal{L}(V,W)$. We have

$$\begin{aligned}
\langle (S+T)^*(w), v\rangle &= \langle w, (S+T)(v)\rangle && \text{(definition of adjoint)}\\
&= \langle w, S(v)+T(v)\rangle && \text{(definition of } S+T)\\
&= \langle w, S(v)\rangle + \langle w, T(v)\rangle && \text{(linearity of } \langle\cdot\rangle)\\
&= \langle S^*(w), v\rangle + \langle T^*(w), v\rangle \\
&= \overline{\langle v, S^*(w)\rangle} + \overline{\langle v, T^*(w)\rangle} && \text{(conjugate symmetry)}\\
&= \overline{\langle v, S^*(w)\rangle + \langle v, T^*(w)\rangle} && \text{(property of conjugate)}\\
&= \overline{\langle v, S^*(w)+T^*(w)\rangle} && \text{(linearity of } \langle\cdot\rangle)\\
&= \langle S^*(w)+T^*(w), v\rangle && \text{(conjugate symmetry)}
\end{aligned}$$

Since the above holds for all $w \in W$, we have $(S+T)^*(w) = S^*(w) + T^*(w)$.

II) Let $\alpha \in \mathbb{C}$. Again, note $\alpha T \in \mathcal{L}(V,W)$ since $\mathcal{L}(V,W)$ is a vector space. Hence, it makes sense to define the adjoint operator of αT.

$$\begin{aligned}
\langle (\alpha T)^*(w), v\rangle &= \langle w, (\alpha T)(v)\rangle && \text{(definition of adjoint of } (\alpha T)^*)\\
&= \langle w, \alpha T(v)\rangle && \text{(definition of } \alpha T)\\
&= \alpha\langle w, T(v)\rangle && \text{(linearity of } \langle\cdot\rangle)\\
&= \alpha\langle T^*(w), v\rangle && \text{(definition of adjoint)}\\
&= \alpha\ \overline{\langle v, T^*(w)\rangle} && \text{(conjugate symmetry)}\\
&= \overline{\overline{a}\langle v, T^*(w)\rangle} && \text{(property of conjugate)}\\
&= \overline{\langle v, \overline{a}T^*(w)\rangle} && \text{(linearity of } \langle\cdot\rangle)\\
&= \langle \overline{a}T^*(w), v\rangle && \text{(conjugate symmetry)}
\end{aligned}$$

Thus, $(\alpha T)^* = \overline{\alpha}T^*$.

III) Let $T \in \mathcal{L}(V,W)$. We know $T^* \in \mathcal{L}(W,V)$, and $(T^*)^* \in \mathcal{L}(V,W)$. Fix $w \in W$ and let $v \in V$.

$$\begin{aligned}
\langle (T^*)^*(v), w\rangle &= \langle v, T^*(w)\rangle && \text{(definition of adjoint of } (T^*)^*)\\
&= \overline{\langle T^*(w), v\rangle} && \text{(conjugate symmetry)}\\
&= \overline{\langle w, T(v)\rangle} && \text{(definition of adjoint)}\\
&= \langle T(v), w\rangle && \text{(conjugate symmetry)}
\end{aligned}$$

Therefore, $(T^*)^* = T$. ∎

Exercise 5.1.2. Prove Proposition 5.1.1.

Remark 5.1.5. In the above proof, to be able to bring together/break up sums in the first slot of the inner product, we have to apply the conjugate symmetry property to move everything to the second slot and apply linearity. Things would be much nicer if we have some properties similar to linearity that we can use directly in the first slot. It turns out that there is indeed some properties, captured by the following Proposition.

Proposition 5.1.2. *Let $\langle \cdot \rangle$ be an inner product on V over $\mathbb{F}$. Then for all $u, v, w \in V$ and $\alpha \in F$, we have*

I) $\langle u + w, v\rangle = \langle u, v\rangle + \langle w, v\rangle$

II) $\langle \alpha u, v\rangle = \overline{\alpha}\langle u, v\rangle$

The above properties are called the ***antilinearity*** *of $\langle \cdot \rangle$ in the first slot.*

Exercise 5.1.3. Prove Proposition 5.1.2.

Now let us establish the relationship between the subspaces kernel and image of linear maps T and its corresponding adjoint operator T^*.

Theorem 5.1.2. *Suppose $T \in \mathcal{L}(V, W)$. Then*

I) $\ker(T^*) = (\text{rang}(T))^{\perp}$;

II) $\text{rang}(T^*) = (\ker T)^{\perp}$;

III) $\ker(T) = (\text{rang}(T^*)^{\perp}$;

IV) $\text{rang}(T) = (\ker(T^*))^{\perp}$.

Proof. We prove II). You can prove the remaining relationships in a similar manner. Fix $u \in \text{rang}(T^*)$ and take any $v \in \ker(T)$. Since u is in the range of T^* there exists a w in W such that $T^*(w) = u$. Now we have that $\langle u, v\rangle = \langle T^*(w), v\rangle = \langle w, T(v)\rangle = \langle w, \mathbf{0}_W\rangle = 0$. Since u is orthogonal to any vector in $\ker(T)$, we have established the equality between these two subspaces. ■

Theorem 5.1.3. *Let $T \in \mathcal{L}(V, W)$, and α, β be orthonormal bases for V and W respectively. Then . In other words, If $R_{\alpha\beta}$ is the matrix representation of T, then the matrix representation of T^* is the conjugate transpose of the matrix representation of T, $R^*_{\alpha\beta}$.*

Proof. Let $\alpha = \langle q_1, \cdots, q_n \rangle$ and $\beta = \langle e_1, \cdots, e_m \rangle$ be two ordered orthonormal bases for V and W respectively. It is sufficient to show that the entry in the j^{th} column, k^{th} row in the matrix representation of T^* would be the complex conjugate of the entry in the k^{th} column, and j^{th} row of the matrix representation $R_{\alpha\beta}$. Take any $1 \leq j \leq n$ then $T(q_j) = \langle e_1, T(q_j) \rangle e_1 + \cdots + \langle e_m, T(q_j) \rangle e_m$. So the j^{th} column of $R_{\alpha\beta}$ is

$$\left[T(q_j)\right]_\beta = \begin{bmatrix} \langle e_1, T(q_j) \rangle \\ | \\ \langle e_k, T(q_j) \rangle \\ | \\ \langle e_m, T(q_j) \rangle \end{bmatrix}.$$

Similarly, for any $1 \leq j \leq m$ we have that $T^*(e_i) = \langle q_1, T^*(e_j) \rangle q_1 + \cdots + \langle q_n, T^*(e_j) \rangle q_n$. But note that the entry in the j^{th}, column k^{th} row of the matrix representation of T^* is $\langle q_k, T^*(e_j) \rangle = \overline{\langle T^*(e_j), q_k \rangle} = \overline{\langle e_j, T(q_k) \rangle}$, and $\overline{\langle e_j, T(q_k) \rangle}$ is the complex conjugate of the entry in the k^{th} column, and j^{th} row of the matrix representation $R_{\alpha\beta}$ of T.
$\therefore$ The matrix representation of T^* is simply the conjugate transpose of the matrix representation of T, $R_{\alpha\beta}$ ■

Remark 5.1.6. The above theorem fails if the bases are not orthonormal (see example below). Thus, it's much easier to work with operators than with their matrix representations since the former is not dependent on the basis.

Example 5.1.1. Consider the inner product space $\mathbb{P}_2[x]$ where $\langle p(x), q(x) \rangle = \int_0^1 p(x)q(x)dx$. In addition consider the linear map $T(ax^2 + bx + c) = bx$. Let $\langle 1, x, x^2 \rangle$ be an ordered basis of $\mathbb{P}_2[x]$. Then $R_{\mathcal{S}} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} = R^*_{\mathcal{S}}$. But $T \neq T^*$. Take $p(x) = 1$ and $q(x) = x$. Then clearly we have that
$\langle p(x), T(q(x)) \rangle = \langle 1, x \rangle = \frac{1}{2} \neq 0 = \langle 0, x \rangle = \langle T(p(x)), q(x) \rangle$. Hence, T cannot be self-adjoint. The reason that Theorem 5.1.3 is not applicable here is that our standard basis $\mathcal{S}$ is not orthogonal, in particular orthonormal.

Definition 5.1.3 (Self-adjoint Operator)**.** Let V be an inner product space. An **operator** $T \in \mathcal{L}(V)$ is called **self-adjoint** if $T = T^*$. That is $T \in \mathcal{L}(V)$ is self-adjoint if and only if

$$\langle w, T(v) \rangle = \langle T(w), v \rangle$$

for all $v, w \in V$.

Remark 5.1.7. By operators on V, we mean linear maps from V to V.

Lemma 5.1.1. *Suppose V is a* ***complex*** *inner product space and $T \in \mathcal{L}(V)$. If*

$$\langle v, T(v)\rangle = 0 \ \forall v \in V,$$

then $T = 0$.

Proof. Let $u, v \in V$. Since it is not given that $T = T^*$ and the inner product is a function into complex, we can re-write
$\langle u, T(v)\rangle$ as $\frac{\langle u+v, T(u+v)\rangle}{4} - \frac{\langle u-v, T(u-v)\rangle}{4} - (\frac{\langle u+iv, T(u+iv)\rangle}{4} - \frac{\langle u-iv, T(u-iv)\rangle}{4})i$ (1)
But note that by our hypothesis every term in (1) is equal to 0. Hence, $\langle u, T(v)\rangle = 0$ ■

Remark 5.1.8. The above lemma only holds if V is a **complex** inner product space! The following is a counterexample to the theorem in a real inner product space.

Example 5.1.2. Rotation by 90 degrees. Then clearly $\langle T((1,0)), (1,0)\rangle = \langle (0,1), (1,0)\rangle = 0$ and $\langle T((0,1)), (0,1)\rangle = \langle (1,0), (0,1)\rangle = 0$

Corollary 5.1.3.1. *Let V be any inner product space. If $T \in \mathcal{L}(V)$ is a self-adjoint and $\langle v, T(v)\rangle = 0$ for all $v \in V$, then $T = 0$.*

Proof. Since we have already covered the complex inner product space, without loss of generality assume V is a real inner product space. So, $\langle , \rangle : V \times V \to \mathbb{R}$. Since $T = T^*$ we can reduce the expression (1) in the proof of Lemma 5.1.1 to $\frac{\langle u+v, T(u+v)\rangle}{4} - \frac{\langle u-v, T(u-v)\rangle}{4}$ ■

Exercise 5.1.4. Finish the proof of Corollary 5.1.3.1.

Remark 5.1.9. In the above corollary, we relax the restriction of V being complex, since T is self-adjoint. Note that in example 5.1.2, T is not self-adjoint.

It is tedious to verify if an operator T is self-adjoint by explicitly finding T and T^*. The following theorem will help us characterizing self-adjoint operators.

Theorem 5.1.4 (Characterization of self-adjoint operators)**.** *Let V be a complex inner product space and $T \in \mathcal{L}(V)$. Then T is self-adjoint if and only if*

$$\langle v, T(v)\rangle \in \mathbb{R} \quad \forall v \in V$$

Proof.

($\Longrightarrow$) Suppose T is self-adjoint. Let $v \in V$. Then by definition of self-adjoint,

$$\langle v, T(v)\rangle = \langle T(v), v\rangle = \overline{\langle v, T(v)\rangle}$$

Then $\langle v, T(v)\rangle$ is its own complex conjugate, so it must be real.

($\Longleftarrow$) Since $\langle v, T(v)\rangle \in \mathbb{R}$ we have that

$$\begin{aligned}
&\langle v, T(v)\rangle = \overline{\langle v, T(v)\rangle} \\
&\langle T^*(v), v\rangle = \overline{\langle v, T(v)\rangle} \\
&\langle T^*(v), v\rangle - \langle T(v), v\rangle = 0 \\
&\langle T^*(v) - T(v), v\rangle = 0 \\
&\langle (T^* - T)(v), v\rangle = 0 \\
&T^* - T = 0 && \text{(since } v \text{ was arbitrary by Lemma 5.1.1)}
\end{aligned}$$

We leave it as an exercise.

■

Remark 5.1.10. Since self-adjoint operators play an analogous role as that of Hermitian matrices, we can easily conclude that the eigenvalues of a self-adjoint operators are also **REAL**.

Exercise 5.1.5. Prove Remark 5.1.10 as follows: Prove that the eigenvalues of the adjoint operator T^* of T is the complex conjugate of the eigenvalue λ associated with the linear map T. Hence, if $T = T^*$ then we must have that $\lambda = \overline{\lambda} \iff \lambda \in \mathbb{R}$

Remark 5.1.11. Even though we can conclude something about eigenvalues of an adjoint operator, we really cannot conclude anything special about the relationships between the eigenvectors of a linear map and its adjoint, since the domain of each map may be completely different vector space. However, if T and T^* are both defined on the same vector space V and satisfy certain relationships, then **both of these maps yield the SAME eigenvectors, as we shall see below.**

Definition 5.1.4 (Normal Operators)**.** Let V be an inner product space. An operator $T \in \mathcal{L}(V)$ is called **normal** if

$$T^*T = TT^*.$$

The following theorem allows us to characterize normal operators.

Theorem 5.1.5 (Characterizing normal operators). *$T \in \mathcal{L}(V)$ is normal if and only if*

$$\|T(v)\| = \|T^*(v)\| \quad \forall v \in V.$$

Proof.

($\Longrightarrow$) Suppose T is normal. Let $v \in V$. We have

$$\begin{aligned}
T^*T = TT^* &\implies T^*T - TT^* = 0 \\
&\implies \langle v, (T^*T - TT^*)(v)\rangle = 0 \\
&\implies \langle v, (T^*T)(v) - (TT^*)(v)\rangle = 0 \\
&\implies \langle v, (T^*T)(v)\rangle - \langle v, (TT^*)(v)\rangle = 0 && \text{(linearity of } \langle \cdot \rangle) \\
&\implies \langle v, (T^*T)(v)\rangle = \langle v, (TT^*)(v)\rangle \\
&\implies \langle v, T^*(T(v))\rangle = \langle v, T(T^*(v))\rangle && \text{(definition of } T^*T \text{ and } TT^*) \\
&\implies \langle T(v), T(v)\rangle = \langle T^*(v), T^*(v)\rangle && \text{(apply def of adjoint to both sides)} \\
&\implies \|T(v)\|^2 = \|T^*(v)\|^2 && \text{(definition of induced norm)} \\
&\implies \|T(v)\| = \|T^*(v)\|
\end{aligned}$$

($\Longleftarrow$) We leave it as an exercise.

■

Exercise 5.1.6. Prove ($\Longleftarrow$) direction of Theorem 5.1.5.

Now we can also identify an important relationship between the kernel of an operator T and its adjoint if the linear map T is normal:

Corollary 5.1.5.1. *Let V be an inner product space. If T is normal, then* $\ker(T) = \ker(T^*)$.

Proof. This an easy consequence of Theorem 5.1.5 which states normal operators preserve the length of the images of a given vector under both the map and its adjoint. Or more precisely let $\overline{v \in V \text{ such that}}$ $\underline{v \in \ker(T).}$ Note that $\|T(v)^*\|^2 = \langle T^*(v), T^*(v)\rangle = \langle v, T(T^*(v))\rangle = \langle v, T^*(T(v))\rangle = \overline{\langle T^*(T(v)), v\rangle} = \langle T(v), T(v)\rangle = \overline{\|T(v)\|^2} = \|T(v)\|^2 = 0$ since $T(v) = \mathbf{0}$. Hence, we have that $\|T(v)^*\|^2 = 0 \iff v \in \ker(T^*)$ ■

Remark 5.1.12. Note that the converse of Corollary 5.1.5.1 is not true.

Theorem 5.1.6. *Let V be an inner product space and $T \in \mathcal{L}(V)$ be normal. If $v \in V$ is an eigenvector of T corresponding to the eigenvalue λ, then v is also an eigenvector of T^* corresponding to the eigenvalue $\overline{\lambda}$.*

Proof. First note if T is normal, then $T - \lambda I$ is also normal because

$$\begin{aligned}
(T-\lambda I)(T-\lambda I)^* &= (T-\lambda I)(T^* - (\lambda I)^*) \\
&= TT^* - (\lambda I)T^* - T(\lambda I)^* + (\lambda I)(\lambda I)^* \\
&= T^*T - (\lambda I)T^* - T(\lambda I)^* + (\lambda I)(\lambda I)^* && (T \text{ is normal}) \\
&= T^*T - (\lambda I)T^* - T(\lambda I)^* + (\lambda I)^*(\lambda I) \\
& && ((\lambda I)^* \text{ and } (\lambda I) \text{ are diagonal, so they commute}) \\
&= T^*T - T^*(\lambda I) - (\lambda I)^*T + (\lambda I)^*(\lambda I) \\
& && ((\lambda I)T^* = T^*(\lambda I) \text{ and } T(\lambda I)^* = (\lambda I)^*T, \text{ you should check}) \\
&= (T-\lambda I)^*(T-\lambda I)
\end{aligned}$$

Let $v \in \ker(T - \lambda I)$. Then v is an eigenvector of T. We have

$$(T-\lambda I)(v) = 0 \implies \|(T-\lambda I)(v)\| = 0$$

Since $T - \lambda I$ is normal, by Theorem 5.1.5 we have

$$\begin{aligned}
0 &= \|(T-\lambda I)(v)\| \\
&= \|(T-\lambda I)^*(v)\| && \text{(Theorem 5.1.5)} \\
&= \left\|(T^* - \overline{\lambda} I)(v)\right\| && \text{(properties of adjoint)} \\
&\implies (T^* - \overline{\lambda} I)(v) = 0 \\
&\implies v \in \ker(T^* - \overline{\lambda} I)
\end{aligned}$$

Thus, v is an eigenvector of T^* corresponding to the eigenvalue $\overline{\lambda}$. ■

Corollary 5.1.6.1. *Let V be a inner product space and $T \in \mathcal{L}(V)$ be normal. Then eigenvectors of T corresponding to distinct eigenvalues are orthogonal.*

Proof. Let v, u be two eigenvectors of T with corresponding eigenvalues λ_1 and λ_2 respectively. Then $\lambda_2\langle v, u\rangle = \langle v, \lambda_2 u\rangle = \langle v, T(u)\rangle = \langle T^*(v), u\rangle = \langle \overline{\lambda_1} v, u\rangle = \lambda_1\langle v, u\rangle$
Hence, $\underbrace{(\lambda_1 - \lambda_2)}_{\in\mathbb{C}}\underbrace{\langle v, u\rangle}_{\in\mathbb{C}} = 0$. But since $\mathbb{C}$ is a field and $\lambda_1 - \lambda_2 \neq 0$ we get that the eigenvectors u and v are orthogonal. ■

Exercise 5.1.7. Can we strengthen the Theorem 4.3.7 to A being normal since normal is weaker than Hermitian?

Example 5.1.3. Let V be a complex inner product space and $T \in \mathcal{L}(V)$ be a normal operator. Suppose T has eigenvalues $3, 4$ then show there exists a vector v with norm $\sqrt{2}$ and $\|T(v)\| = 5$.

ANSWER: Let u, w be the eigenvectors associated with eigenvalues 3 and 4. Now let $v = \frac{u}{\|u\|} + \frac{w}{\|w\|}$ and clearly the norm of v is $\sqrt{2}$ Why? Use the fact that eigenvectors of a **normal** map corresponding to distinct eigenvalues are orthogonal and then apply **Pythagorean Theorem**. Now $\|T(v)\|^2 = \langle \frac{3u}{\|u\|} + \frac{4w}{\|w\|}, \frac{3u}{\|u\|} + \frac{4w}{\|w\|} \rangle =$ 25